AICT 2013

The Conference is devoted to the 150 anniversary of the Latvia University of Agriculture

LATVIA UNIVERSITY OF AGRICULTURE
FACULTY OF INFORMATION TECHNOLOGIES

APPLIED INFORMATION AND COMMUNICATION TECHNOLOGIES

PROCEEDINGS
OF THE 6-th INTERNATIONAL SCIENTIFIC CONFERENCE
ISSN 2255-8586
JELGAVA, LATVIA, APRIL 25-26, 2013

JELGAVA 2013

Scientific advisory board

Chairman:

Assoc. Prof., Dr.sc.ing. Egils Stalidzans, Latvia University of Agriculture, Latvia

Members:

Prof., Dr.sc.ing. Irina Arhipova, Latvia University of Agriculture, Latvia

Prof., Dr.sc.phys. Uldis Iljins, Latvia University of Agriculture, Latvia

Prof., Dr.hab.sc.ing. Peteris Rivza, Latvia University of Agriculture, Latvia

Prof., Dr.sc.paed. Anda Zeidmane, Latvia University of Agriculture, Latvia

Assoc. Prof., Dr.sc.comp. Rudite Cevere, Latvia University of Agriculture, Latvia

Assoc. Prof., Dr.sc.agr. Liga Paura, Latvia University of Agriculture, Latvia

Assoc. Prof., Dr.sc.phys. Uldis Gross, Latvia University of Agriculture, Latvia

Assoc. Prof., Dr.sc.oec. Aleksandrs Gailums, Latvia University of Agriculture, Latvia

Assist. Prof., Dr.sc.oec. Liga Ramute, Latvia University of Agriculture, Latvia

Assist. Prof., Dr.sc.paed. Larisa Turuseva, Latvia University of Agriculture, Latvia

Assist. Prof., Dr.sc.paed. Nataļja Vronska, Latvia University of Agriculture, Latvia

Assist. Prof., Dr.sc.ing. Gatis Vītols, Latvia University of Agriculture, Latvia

Prof., Dr.hab.sc.ing. Jānis Grundspeņķis, Riga Technical University, Latvia

Prof., Dr.hab.sc.ing. Jurijs Merkurjevs, Riga Technical University, Latvia

Assist.prof., Dr.sc.ing. Agris Nikitenko, Riga Technical University, Latvia

Prof., Dr.sc.comp. Leo Selavo, University of Latvia, Latvia

Prof., Dr.sc.comp. Juris Vīksna, University of Latvia, Latvia

Prof., Dr.sc.ing. Egils Ginters, Vidzeme University College, Latvia

Assist. Prof., Dr.sc.ing. Arnis Cirulis, Vidzeme University College, Latvia

Assoc. prof., Dr.sc.ing. Pēteris Grabusts, Rezekne Higher Education Institution, Latvia

Assoc. prof., Dr.sc.comp. Anita Jansone, Liepaja Univserity, Latvia

Assoc.prof., Dr.math. Galina Hilkevica, Ventspils University College, Latvia

Dr.sc.ing. Juris Vanags, Association of Biotechnology of Latvia, Latvia

Dr.sc.ing. Ivars Mozga, SIA TIBIT, Latvia

Prof., Dr.sc.oec. Povilas Domeika, Aleksandras Stulginskis University, Lithuania

Prof. Dr.sc.ing. Vytautas Galvanauskas, Kaunas University of Technology, Lithuania

Prof. Dr. Benjamim Fonseca, UTAD University, Portugal

Prof. PhD Pavel Nastase, Bucharest Academy of Economic Studies, Romania

Assoc. prof. Dr.sc.ing. Juozas Navickas, Aleksandras Stulginskis University, Lithuania

PhD Julio Vera González, University of Rostock, Germany

Dr.sc.biol. Piotr Skubida, Research Institute of Pomology and Floriculture, Poland

Prof., RNDr. Jaroslav Pokorny, Charles University in Prague, Czech Republic

Organizing Commitee

Chairman:

Assoc. Prof., Dr.sc.oec. Aleksandrs Gailums, Latvia University of Agriculture, Latvia

Members:

Assoc. Prof. Egils Stalidzans, Latvia University of Agriculture

Mg.sc.ing. Ilona Odzina, Latvia University of Agriculture

Mg. math. Anna Vintere, Latvia University of Agriculture

Mg.sc.ing. Dace Zutere, Latvia University of Agriculture

Mg.paed. Olga Civzele, Latvia University of Agriculture

Dr.sc.ing. Ivars Mozga, Latvia University of Agriculture

B.sc.ing. Nauris Akmenlauks, Latvia University of Agriculture

The authors are responsible for the orthography of submitted materials.

APPLIED INFORMATION AND COMMUNICATION TECHNOLOGIES 2013

The Conference is devoted to the 150 anniversary of the Latvia University of Agriculture

Publisher: Faculty of Information Technologies, Latvia University of Agriculture, Liela iela 2, Jelgava, LV-3001, Latvia
Supported by:.

ccsystems.lv

Centre of Technology Transfer (from LUA)

Photo on Cover: Palace of Jelgava
Author: Marcin Szala

CONTENTS

FOREWORD

The 6-th Scientific Conference APPLIED INFORMATION AND COMMUNICATION TECHNOLOGIES (AICT 2013) is devoted to the 150 anniversary of the Latvia University of Agriculture. Faculty of Information Technologies of Latvia University of Agriculture was founded in 2001, but scientific activities related to information technologies began in the middle of 1960s.

Conference AICT 2013 aims to cover practical aspects of ICT applications in different fields encouraging interdisciplinary approach and contacts among scientists from different branches. Our goal is to bring together researchers, scientists, engineers, practitioners, educators and students to achieve.

The collection of international scientific conference proceedings contains 45 reports by authors from Latvia, Lithuania, Russian Federation, Canada, Czech Republic, Romania, Poland, Republic of Kazakhstan. The reports are divided into the following seven parts:

- Modelling and Simulation Technologies,
- ICT Tools and Methods,
- ICT for Rural development and Agriculture,
- Industrial Applications of ICT,
- WEB Technologies,
- ICT in Education,
- Software development.

We are sure that because of the high quality of the participants the goals of the conference will be met. Many thanks to all the contributors and participants in the conference.

Assoc. prof. Aleksandrs Gailums

Chairman of Organizing Committee
Latvia University of Agriculture

ICT TOOLS AND METHODS

Analysis of ontology based approach for clustering tasks

Peter Grabusts

Engineering Faculty, Rezekne Higher Educational Institution, Atbrivoshanas alley 90, Rezekne, LV-4601, Latvia
peter@ru.lv

Abstract: *Clustering algorithms are used to group given objects defined by a set of numerical properties in such a way that the objects within a group are more similar than the objects in different groups. All clustering algorithms have common parameters the choice of which characterizes the effectiveness of clustering. The most important parameters characterizing clustering are: metrics, number of clusters k and cluster validity criteria. In classical clustering algorithms semantic knowledge is ignored. This creates difficulties in interpreting the results of clustering. Currently, the possibility to use ontology opportunities is developing rapidly, that provide an explicit model for structuring concepts, together with their interrelationship, that allows to gain knowledge on a specific data model. According to the previously obtained results of clustering study, the author will make a first attempt to create ontology based prototype of clustering concepts from numerical data using similarity measures, cluster numbers, cluster validity and others characteristic features.*

Keywords: clustering, cluster analysis, ontology.

Introduction

Nowadays there is a large amount of data accumulated in the various fields of science, business, economy, etc. areas and there is a need to analyze them for better management of one separate sector. Often business requirements stimulate the development of new intelligent data analysis methods that are focused on practical applications. Clustering is as one of the intelligent data analysis tasks and its aim is to search for an independent group (cluster) and its performance in the test data (Everitt et al., 1993). Resolving such a task leads to better understand of the data, because clustering can be used practically in any area of application where data analysis is required.

Author's research interests have been oriented to clustering analysis: clustering algorithms, fuzzy clustering, rule extraction from clustered data etc (Crawen et al., 1994; Hoppner et al., 1999). The next step in the research would be the implementation of ontologies in cluster analysis.

To evaluate the clustering performance aspects the following purpose was put forward – to analyze and summarize the clustering algorithms possibilities in order to create an ontological prototype for numerical data clustering. The work being carried out and following tasks were set:

- to carry out the evaluation of the validity of metrics chose;
- characterize the change in the number of clusters of analyzed data;
- evaluate the reliability of the results of clustering (clusters validity);
- extract the rules from the clusters.

According to the previously obtained results of clustering study, the author will make an attempt to create ontology based prototype of clustering concepts using similarity measures, cluster numbers, cluster validity and others characteristic features.

Clustering tasks and knowledge extraction

Clustering is based on the hypothesis of compactness. It is assumed that the set of elements of the learning characteristics of the space is in a compact way. The main task is a formalized description of these formations. Some clustering algorithms are well known, such as Isodata, FOREL, k-means etc.

Clustering differs from classification in following - in cluster analysis process there is no need to distribute a separate variable group. From this point of view, clustering is considered as a "learning without a teacher" and is used in the initial stage of the research.

Clustering is characterized by two features that distinguish it from other methods:

- the result depends on the object itself or the kind of attributes used, namely, they can be clearly defined objects or objects with fuzzy description;
- the result depends on the potential of the cluster and the object relations of natural clusters, that is, we should take into account the possible object belonging to multiple clusters and object ownership detection (strong or fuzzy membership).

Given the important role of clustering in data analysis, object ownership concept was generalized to a class function that defines the belonging of class object to proper class. Two classes of characteristic functions are distinguished:

- a discrete function that accepts one of two possible values - belong / not belong (classical clustering);

- a function that accepts values from the interval [0,1]. The closer the value of function is to 1, the more the subject belongs to a certain class (fuzzy clustering).

Clustering algorithms are mainly designed for multi-dimensional data sampling processing, when the data are given in tabular form in the "object-property". They allow you to group objects into groups, where objects relate to each other by a specific rule. It does not matter, how these groups are named- taxons, clusters, classes, as long as it reasonably reflects the properties of this object. After clustering the data for further analysis other intelligent data analysis techniques are used to determine the nature of the resulting regularities and future uses.

Clustering is typically used for data processing as a first step in the analysis. It identifies groups of similar data that can later be used for the investigation of relationships between the data. Clustering formal process consists of the following stages:

- collecting the required data for analysis;
- determination the characteristics of the clusters;
- data grouping in clusters;
- class hierarchy definition and analysis of the results.

All clustering algorithms have common characteristics, the selection of which is characterized by a clustering efficiency. The most important clustering parameters are as follows: metric (cluster element distance to the cluster center), the number of clusters k, clustering validity assessment, opportunity to get rules (Gan et al., 2007; Kaufman et al., 2005).

Metrics. The main purpose of metrics learning in a specific problem is to learn an appropriate distance/similarity function. A metrics or distance function is a function which defines a distance between elements of a set (Li et al., 2004; Vitanyi, 2005). A set with a metric is called a metric space. In many data retrieval and data mining applications, such as clustering, measuring similarity between objects has become an important part. In general, the task is to define a function Sim(X,Y), where X and Y are two objects or sets of a certain class, and the value of the function represents the degree of "similarity" between the two. Formally, a distance is a function D with nonnegative real values, defined on the Cartesian product X x X of a set X. It is called a metrics on X if for every x,y,z ϵ X:

- D(x,y)=0 if x=y (the identity axiom);
- D(x,y) + D(y,z) ≥ D(x,z) (the triangle inequality);
- D(x,y)=D(y,x) (the symmetry axiom).

A set X provided with a metric is called a metric space.

Euclidean distance is the most common use of distance – it computes the root of square differences between coordinates of a pair of objects:

$$D_{XY} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \tag{1}$$

Manhattan distance or city block distance represents distance between points in a city road grid. It computes the absolute differences between coordinates of a pair of objects:

$$D_{XY} = \sum_{k=1}^{d} | x_{ik} - x_{jk} | \tag{2}$$

Minkowski distance is the generalized metric distance:

$$D_{XY} = (\sum_{k=1}^{d} | x_{ik} - x_{jk} |^p)^{1/p} \tag{3}$$

Note that when p=2, the distance becomes the Euclidean distance. When p=1, it becomes city block distance.

Cosine distance is the angular difference between two vectors:

$$D_{XY} = \cos(\theta) = \frac{X \bullet Y}{\|X\|\|Y\|} = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n} (X_i)^2} \times \sqrt{\sum_{i=1}^{n} (Y_i)^2}} \tag{4}$$

The summary of the metrics is shown in the Table 1.

Table 1

Distance measures and their applications

Measure	Examples and applications
Euclidean distance	K-means with its variations
Manhattan distance	Fuzzy ART, clustering algorithms
Cosine distance	Text Mining, document clustering

Traditionally Euclidean distance is used in the clustering algorithms, the choice of other metric in definite cases may be disputable. It depends on the task, the amount of data and on the complexity of the task.

Cluster numbers. An important issue in the implementation of clustering algorithm is the number of clusters and initial centers determination. In simple tasks it is assumed that *a priori* is known the number of clusters and as the initial cluster centers m values is offered to take the first set of training points m.

Clustering validity. Cluster validity is a method to find a set of clusters that best fits natural partitions (number of clusters) without any class information. There are three fundamental criteria to investigate the cluster validity: external criteria, internal criteria, and relative criteria (Xu et al., 2009). In this case only external cluster validity index was analyzed.

Given a data set X and a clustering structure C derived from the application of a certain clustering algorithm on X, external criteria compare the obtained clustering structure C to a pre-specified structure, which reflects a priori information on the clustering structure of X. For example, an external criterion can be used to examine the match between the cluster labels with the category labels based on *apriori* information.

Based on the external criteria, there is the following approach: comparing the resulting clustering structure C to an independent partition of the data P, which was built according to intuition about the clustering structure of the data set.

If P is a pre-specified partition of data set X with N data points and is independent of the clustering structure C resulting from a clustering algorithm, then the evaluation of C by external criteria is achieved by comparing C to P. Considering a pair of data points x_i and x_j of X, there are four different cases based on how x_i and x_j are placed in C and P.

- Case 1: xi and xj belong to the same clusters of C and the same category of P.
- Case 2: xi and xj belong to the same clusters of C but different categories of P.
- Case 3: xi and xj belong to different clusters of C but the same category of P.
- Case 4: xi and xj belong to different clusters of C and different category of P.

Correspondingly, the numbers of pairs of points for the four cases are denoted as a, b, c and d. Because the total number of pairs of points is N(N-1)/2, denoted as M, we have:

$$M = a + b + c + d = \frac{n(n-1)}{2} \tag{5}$$

where n is the number of data points in the data set. When C and P are defined, one can choose one of the many clustering quality criteria. In the given research clustering quality criteria have been evaluated with the help of Rand or Hubert index (Xu et al., 2009).

Rand index is calculated by using the following formula:

$$R = \frac{a+d}{M} \tag{6}$$

Rand index suggests an objective criterion for comparing two arbitrary clusterings based on how pairs of data points are clustered. Given two clusterings, for any two data points there are two cases:

- The first case is that the two points are placed together in a cluster in each of two clusterings or they are assigned to different clusters in both clusterings.
- The second case is that the two points are placed together in a cluster in one clustering and they are assigned to different clusters in the other.

Hubert index is calculated by using the following formula:

$$\Gamma = \frac{1}{M} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} Y_{ij} \tag{7}$$

The value of both index ranges between 0 and 1. A higher index value indicates greater similarity between C and P.

Rule extraction. The possibility of directly converting clustering information in the form of symbolic knowledge extraction is through the rules (rule extraction). These assumptions are formulated as IF ... THEN ... rules (Russel et al., 2010). The benefits of the mining rules are as follows:

- the opportunity to verify the extracted rules on different variants of the input data is given;
- failures of training data can be identified, thus clustering operation can be improved by introducing new or removing additional clusters;
- determination of a previously unknown regularities in the data that currently have a growing importance of Data Mining industry;
- the resulting rules can be set up as a base of rules, which might also be used for similar types of applications.

Several artificial neural network algorithms use clustering during the learning process, leading to hidden units, which are, actually, cluster centers (Hush et al., 1993). The nature of each hidden unit enables a simple translation into a single rule:

$$\text{IF Feature}_1 \text{ is TRUE AND IF Feature}_2 \text{ is TRUE ... AND IF Feature}_n \text{ is TRUE}$$
$$\text{THEN Class}_x \tag{8}$$

where a *Feature* is composed of upper and lower bounds calculated by the center μ_n positions, width σ and feature steepness S. The value of the steepness was discovered empirically to be about 0.6 and is related to the value of the width parameter. The values of μ and σ are determined by the training algorithm. The upper and lower bounds are calculated as follows:

$$X_{lower} = \mu_i - \sigma_i + S \text{ and } X_{upper} = \mu_i + \sigma_i - S \tag{9}$$

Then rule extraction RULEX process can be seen below in Table 2 (Andrews et al., 1995).

Table 2

Rule extraction algorithm

Procedure:
For each hidden unit:
For each μ_i
$X_{lower} = \mu_i - \sigma_i + S$
$X_{upper} = \mu_i + \sigma_i - S$
Build rule by:
antecedent=[X_{lower}, X_{upper}]
Join antecedents with AND
Add class label
Write rule

Consequently, a base for the rules has been obtained.

Ontology based approach

In this paper the author presents a formal clustering ontology framework concept, which can provide the background for numerical data clustering. Using the ontology, numerical clustering can become a knowledge-driven process.

As it was mentioned in the previous chapter, clustering is used at the data level instead of the knowledge level, that helps with identifying targets precisely and understanding the clustering results.

Existing clustering methods consider various constraints and they only consider limited knowledge concerning the domain and the users. In such a way, to include domain knowledge in the clustering methods and clustering process becomes an important topic in clustering data research and analysis.

There are many different definitions of ontology but the most common is recognized the following: an ontology is a formal explicit specification of a shared conceptualization (Gruber, 1993). Ontologies are often equated with taxonomic hierarchies of classes. It can be said, that the aim of ontology is to accumulate knowledge in general and formal way.

Ontologies can be classified into different forms. One of the most popular types of classification is offered by Guarino, who classified types of ontologies according to their level of dependence on a particular task or point of view (Guarino, 1998):

- *Top-level ontologies*: describe general concepts like space, time, event, which are independent of a particular problem or domain.
- *Domain-ontologies*: describe the vocabulary related to a generic domain by specializing the concepts introduced in the top-level ontology.
- *Task ontologies*: describe the vocabulary related to a generic task or activity by specializing the top-level ontologies.

- *Application ontologies*: they are the most specific ones. Concepts often correspond to roles played by domain entities. They have a limited reusability as they depend on the particular scope and requirements of a specific application.

It should be noted that ontologies are widely used in document clustering and Semantic Web but numerical data clustering is undeservedly forgotten.
Thus, an ontology is an explicit representation of knowledge. It is a formal, explicit specification of shared conceptualizations, representing the concepts and their relations that are relevant for a given domain of discourse (Gruber, 1993).
The newly developed numerical data clustering otology concept is composed of the following classes:

Clustering_Task. It is an abstract class. It is related to the proper clustering algorithm class. Depending on the purpose and the clustering area (domain), the clustering algorithm, the number of clusters and the data samples are chosen.

Clustering_Algorithm. This class represents a list of available clustering algorithms and their features (Fig.1).

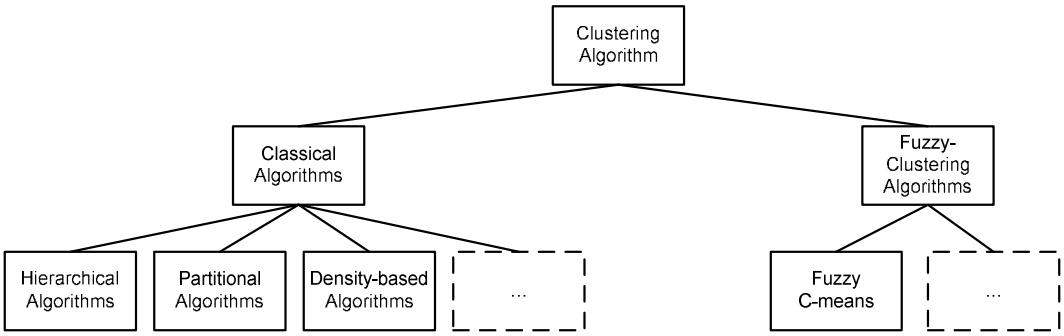


Fig. 1. **A hierarchical view of the *Clustering_Algorithm* class.**

Clustering_Metric. This class represents a list of available distance metrics for clustering algorithms (Fig. 2).

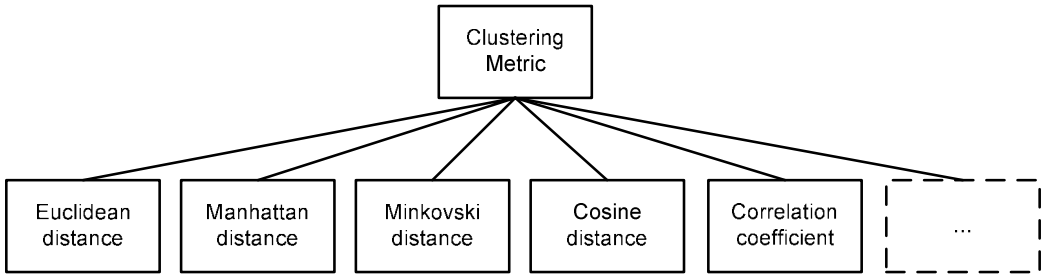


Fig. 2. **A hierarchical view of the *Clustering_Metric* class.**

Clustering_Validity. This class represents a list of cluster validity methods (Fig. 3).

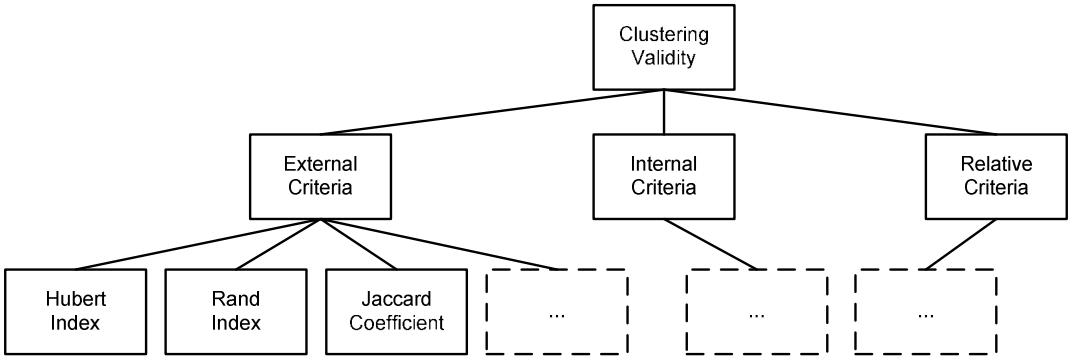


Fig. 3. **A hierarchical view of the *Clustering_Validity* class.**

Clustering_Rule. This class represents a list of rule extraction methods from clusters (if it is possible).

Based on such class analysis the following approach is offered for ontology-based clustering, as shown in Fig. 4.

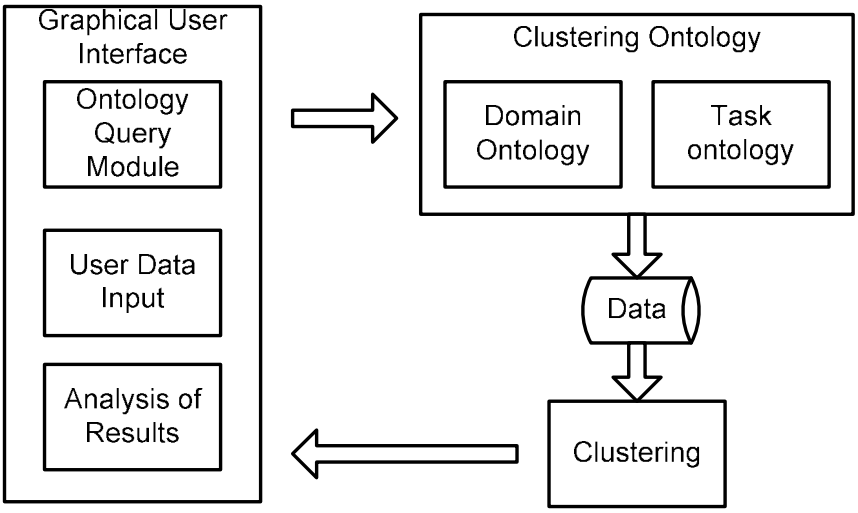


Fig. 4. **The framework concept of ontology-based numerical data clustering.**

Developing framework Protégé OWL tool is used for construct this concept.

Experimental part: bancruptcy data analysis

Clustering ontology prototype should work according to the following scheme: numerical data selection, choice of clustering algorithm, determining the number of clusters, performance of clustering, validation of clustering, acquisition of rules (if possible) (Fig. 5).

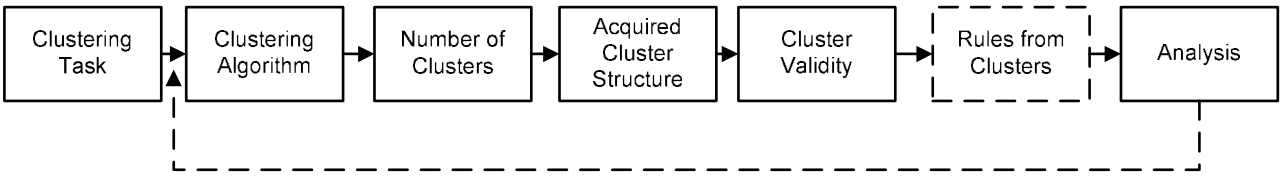


Fig. 5. **Working scheme of clustering prototype.**

The data on firm bankruptcy were taken from (Rudorfer, 1995). For the purpose of experiments, balance sheet data of 63 companies were used (46 - bankruptcy and 17 - not bankruptcy). It was decided to calculate the following financial ratios on the basis of the data available and further use them in all the experiments (Altman, 1968):

- R3: Cash Flow / Total Assets;
- R7: Current Assets / Current Liabilities;
- R9: Current Assets / Total Assets;
- R31: Working capital / Total assets.

The use of clustering algorithm k-means for this data set showed the following results: 59 – bankruptcy and 4 – not bankruptcy.

In order to verify clustering validity, quality index has been calculated – Rand and Hubert index for five clusters. Cluster structure C (consecutively with the number of clusters between 2 and 5 clusters) has been compared with specified divisions P containing various possible clusters.

Further, the total error has been calculated. The following errors of overall clustering have been calculated: for 2 clusters – 66.67 % (error rate of cluster 1: 71.19 %, error rate of cluster 2: 0.00 %).

Among all structures the lowest mistake occurs with 2 clusters, namely, 2 cluster structure in this case is the most optimal. Fig. 6 shows the calculated Rand and Hubert index for 2 cluster structure.

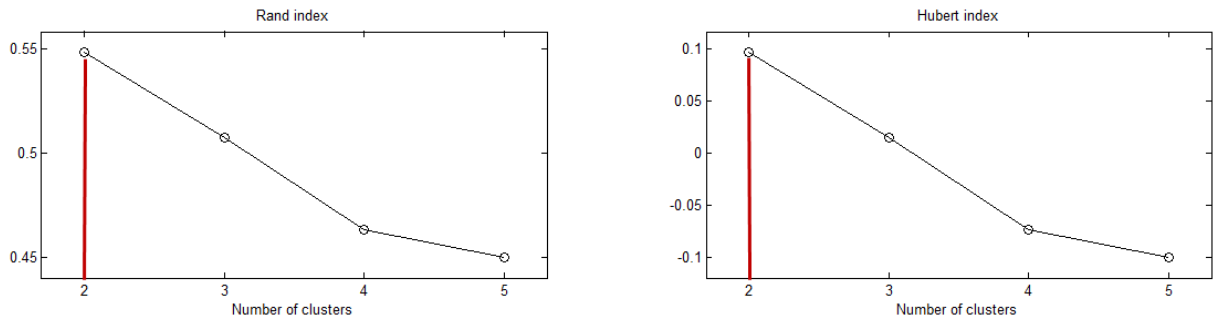


Fig. 6. **Rand and Hubert index**.

It should be noted that for this experiment, the data sample selected is not very successful, but nevertheless it illustrates the principles of clustering algorithm work. Indexes characterizing the quality of clustering are useful for analyzing the performance of clustering algorithms. With their help it is possible to choose an optimal cluster structure in cases when data distribution in clusters has not initially been set.

The objective of the next experiment was to extract rules from the bankruptcy data. Table 3 shows the results obtained in the course of the experiment whereas Table 4 lists the rules obtained under separate S values (Cluster 1 contains data on not bankrupt but Cluster 2 – data on bankrupt companies).

From Table 3 it can be seen that the rules obtained correctly describe bankruptcy data (44 out of 46) within the whole domain of parameter S, i.e., it can be stated that bankruptcy data are located in a fairly compact class.

Table 3

Results of bankruptcy data set training

	Values of parameter S								
Correct	-0.9	…	0	0.1	0.2	0.3	0.4	0.5	0.6
Cluster 1	15		12	10	9	6	5	1	0
Cluster 2	44		44	44	44	44	44	44	44
%	93.7	…	88.9	85.7	84.1	79.4	77.8	71.4	69.8

Table 4

Bankruptcy data set: characteristics of the extracted rules

	Parameter S= -0.9	**Parameter S= 0.4**
Values of centres and radii	Class 1= 0.03 1.25 0.74 0.01 Class 2= 0.13 1.86 0.59 0.10 Values of radii = 0.68 3.92	Class 1= 0.03 1.25 0.74 0.01 Class 2= 0.13 1.86 0.59 0.10 Values of radii = 0.68 3.92
Rules correctly describe elements of classes (%)	93.7	77.8
Rule of Cluster 1	IF (X1>= -1.54 AND < 1.61) AND IF (X2>= -0.33 AND < 2.83) AND IF (X3>= -0.84 AND < 2.32) AND IF (X4>= -1.57 AND < 1.59) THEN NON-BANKRUPT	IF (X1>= -0.24 AND < 0.31) AND IF (X2>= 0.97 AND < 1.53) AND IF (X3>= 0.46 AND < 1.02) AND IF (X4>= -0.27 AND < 0.29) THEN NON-BANKRUPT
Rule of Cluster 2	IF (X1>= -4.69 AND < 4.95) AND IF (X2>= -2.97 AND < 6.68) AND IF (X3>= -4.23 AND < 5.41) AND IF (X4>= -4.72 AND < 4.93) THEN BANKRUPT	IF (X1>= -3.39 AND < 3.65) AND IF (X2>= -1.67 AND < 5.38) AND IF (X3>= -2.93 AND < 4.11) AND IF (X4>= -3.42 AND < 3.63) THEN BANKRUPT

Conclusion

In clustering there is no directly formalized criterion, so different clustering parameters are chosen by subjective assessment. This refers to the selection of clustering algorithm and the number of clusters in each case, also to the cluster validation criteria. Also it is very important to get knowledge from clusters in the form of rules. All this leads to some problems in interpreting the results of clustering. In recent decades cluster analysis has been transformed from one of the data analysis sections in a separate direction, which is closely related to knowledge support system. Partly, this has happened due to the introduction of the ontology concept in the clustering characteristics description. The use of clustering ontology for documents and Semantic Web applications is

expanding rapidly but the numerical data clustering is undeservedly neglected. The author has made an attempt to define and develop an ontology-based prototype for clustering numerical data. This concept contains several concept classes: clustering algorithms, numbers of clusters, cluster validity and other characteristic features. In future studies these classes refinement will be carried out, also the real model according to data clustering purpose will be worked up.

References

Altman, E., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. In:*Journal of Finance*, vol. 13, pp.589-609.

Andrews, R. and Gewa, S., 1995. RULEX and CEBP networks as the basis for a rule refinement system. In: *J. Hallam et al, editor, Hybrid Problems, Hybrid Solutions*. IOS Press.

Crawen, M. and Shavlik, J., 1994. Using sampling and queries to extract rules from trained neural networks. In:*Machine Learning: Proceedings of the Eleventh International Conference,* San Francisco, CA.

Everitt, B.S., 1993. Cluster analysis. *John Viley and Sons*, London, 170 p.

Gan, G., Ma, C., Wu, J., 2007. Data clustering: Theory, algorithms and applications. *ASA-SIAM series on Statistics and Applied Probability*, SIAM, Philadelphia, ASA, Alexandria, VA, 2007.

Gruber, T. R., 1998. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2) (1993) 199-220.

Guarino, N., 1998. Formal Ontology in Information Systems. In:*1st International Conference on Formal Ontology in Information Systems*, FOIS, Trento, Italy, IOS Press, 3-15.

Hush, D.R. and Horne, B.G., 1993. Progress in Supervised Neural Networks. What's new since Lippmann? In: *IEEE Signal Processing Magazine*, vol.10, No 1.,p.8-39.

Hoppner, F., Klawonn, F., Kruse, R., Runkler, T., 1999. Fuzzy Cluster Analysis. *John Whiley and Sons,* New York.

Kaufman, L. and Rousseeuw, P.J., 2005. Finding groups in data. An introduction to cluster analysis. *John Wiley & Sons.*

Li, M., Chen, X., Ma, B., Vitanyi, P., 2004. The similarity metric. In:*IEEE Transactions on Information Theory*, vol.50, No. 12, pp.3250-3264.

Protégé project homepage: http://protege.stanford.edu/index.html , 05.03.2013.

Rudorfer, G., 1995. Early bankruptcy detecting using neural networks. In:*APL Quote Quad*, ACM New York, vol. 25, N. 4, P. 171-178. (Data available at: http://godefroy.sdf-eu.org/apl95/ratios95.zip, 05.03.2013).

Russel, S. and Norvig, P., 2010. Artificial Intelligence: A Modern Approach. *Prentice Hall*, 1132 p.

Vitanyi, P., 2005. Universal similarity. In:*ITW2005*, Rotorua, New Zealand.

Xu, R. and Wunch,D.C., 2009. Clustering. *John Wiley & Sons*, pp. 263-278.

Automatic extraction of geographic context from textual data

Jurijs Nikolajevs, Gints Jekabsons

Institute of Applied Computer Systems, Riga Technical University, Meza street 1/3, Riga, LV-1048, Latvia
juryn@inbox.lv

Abstract: *The amount of information on the internet grows exponentially. It is not enough anymore just to have a general access to this huge amount of data, instead it is becoming a necessity to be able to use different kinds of automatic filters to retrieve just the information you actually want. One solution for the information filtering and retrieval is context analysis in which one of the contexts of interest is the geographic context. This paper studies the problem and methodology of geoparsing – recognition of geographic names in unstructured textual content for the aim of extracting geographic context. A prototype implementation of a geoparsing system, capable of automatically analyzing unstructured text, recognizing geographic information and marking geographic names, is developed. Empirical evaluation of the system using articles from real-world news showed that the average quality of its geographic name recognition varies around 75-100%. Possible applications of the developed prototype include automated grouping of any texts by their geographic contexts (e.g., in news portals) and location-based search. Preliminary results of empirical evaluation showed that the average rate of its geographic name recognition varies around 75-100%.*

Keywords: geoparsing, geocoding, geographic information retrieval, natural language processing.

Introduction

Nowadays, there are a lot of technologies and simple (web) services allowing to manually geotag photos, videos, texts, and other information, but there is still a lack of systems that can extract the geographic context automatically. This study tackles with automatic extraction of geographic context from unstructured text documents.

The great majority of textual data in the web is not directly linked to geographic context. Such linking would be very useful for information searching and structuring. Automated geoparsing (Goldberg, 2008; Keller et al., 2009; Abascal-Mena et al., 2013; Nikolajevs and Jekabsons, 2011) can help solving this problem by efficiently automating the linking of the textual data to geographic identifiers (e.g., codes or geographic coordinates expressed in form of latitude-longitude). These links can then be used in search engines and other applications that can operate with geographic data. The geocoded texts can also be linked with descriptions of the places as well as their photos, videos, etc. Such geoparsed texts can provide reader with additional information about the geographic context of the text. The user can obtain the additional information about geographic objects mentioned in the text by simple click on the objects name in the text. Sets of geoparsed texts can be structured according to the locations mentioned in them allowing their readers to select the most appropriate information in the shortest time (one of the examples could be the news portal that structures sets of articles by their geographic content). Identified geographic places can also be displayed on a geographic map.

There are a number of commercial products with geoparsing capabilities. Companies like MetaCarta extract information about place and time, Digital Reasoning (GeoLocator), Lockheed Martin (AeroText), and SRA (NetOwl) (Caldwell, 2009) extract places along with other entities, such as persons, organizations, time, money, etc. Many of the existing services analyze only their specific fields. For example, HealthMap is part of a new generation of online systems designed to monitor and visualize, on a real-time basis, disease outbreak alerts as reported by online news media and public health sources (Keller et al., 2009).

This paper studies the methodology of geoparsing as well as related problems and solutions thereof, describes main principles and implementation of geoparsing ideas into a prototype, and presents the results of its empirical evaluation.

Methodology

Geoparsing is the process of assigning geographic identifiers (e.g., codes or geographic coordinates expressed as latitude-longitude) to words and phrases that occur in unstructured textual content. It can be divided into two main phases (Goldberg, 2008; Nikolajevs and Jekabsons, 2011): (1) recognition of geographic names in text and (2) assignment of the most probable geographic identifiers to the words designating the real world objects (geocoding).

Phase 1: Recognition of geographic names

To recognize potential geographic names in an unstructured text, the so-called external (or gazetter-based) approach can be used (Goldberg, 2008; Mikheev et al., 1999). This approach requires a gazetteer database that contains geographic names (including their variations) and their locations (geographic coordinates). According

to study in (Mikheev et al., 1999), the external approach gives higher recognition results in comparison with internal approaches which recognize geographic names using only the information given in text.
In simplest case of external approach, each text word that begins with a capital letter is checked for its existence in the gazetteer database. Additionally, the database is also queried for such word combinations where the second word indicates that the previous one is a geographic object (e.g., where the second word is "lake", "river", "bay", etc.) as well as two or more consecutive words that start with capital letters. Also, to be able to recognize old place names or place names in other languages, gazetteer database must contain such alternate names. Additionally, because of linguistic features found in many languages it might be necessary to transform all words to their basic primary form before querying the database (it can be done by analyzing the ending of each word and, according to the grammar rules, returning all possible basic forms of the word or word combination).
The result of this phase is an unfiltered list of (potentially many) geographic names that potentially could refer to actual geographic places mentioned in the text.

Phase 2: Assignment of the most probable geographic identifiers

Previous phase returned a list of potential geographic names from the gazetteer. At this point it is not clear, whether a potential place name is really the place mentioned in the text. Therefore, we need to develop heuristics for ranking of the place names found in gazetteer. But this is not a trivial task. For example, there are also many names of geographic objects, that could are used as regular words in other languages, or are homonymic with person names, e.g., city of Paris and person Paris Hilton. Also many place names are not unique – a number of European cities have duplicates in the "New World" where the settlers named their colonies in honor of their countries or cities in Europe (for example, the city of Newcastle in England and Newcastle in Australia). Street names also can be homonymic to city names. To solve some of these problems, context analysis or special stop-lists can be created. Using context analysis, some other place names found in the text can point to the region where the place is located (e.g., USA or UK in case of Newcastle).
In the developed prototype, each place name is evaluated according to three criteria: First, every name found in the database should be compared to the original name, which was queried to the database. If the returned queried word exactly matches a place name, that word gets the highest rating – the highest level of reliability. If it matches an alternative name of the place it gets medium level of reliability. If there is only a partial match of the names it gets lower level of reliability. Let R_1 be normalized reliability:

$$R_1 = \frac{L_{cur}}{L_{\max}} \tag{1}$$

where L_{cur} – reliability for the current place;
L_{max} – maximal possible reliability.

Let R_2 be normalized frequency of the name occurrence in the text (if the name appears in the text more often, the probability that the text is about this geographic object increases):

$$R_2 = \frac{F_{cur}}{F_{\max}} \tag{2}$$

where F_{cur} – frequency of occurrence for the current place;
F_{max} – maximal frequency of occurrence.

Let R_3 be normalized evaluation of mutual arrangement of the referred geographic objects on geographic map (the whole text is most likely related to the region where most of the found places are situated):

$$R_3 = \frac{\sum_{i=0}^{n} d(a_{cur}, a_i)}{\sum_{i=0}^{n-1} \sum_{k=i+1}^{n} d(a_i, a_k)} \tag{3}$$

where d – distance between two point on the Earth surface;
a – one of the geographic objects found in the text (a_{cur} points to the current object);
n – total amount of geographic objects found in the text.

Now the final rating of the current geographic object is:

$$R_{final} = \frac{1}{3}(R_1 + R_2 + R_3)\,. \tag{4}$$

This evaluation is applied for all potential geographic objects found in the text and finally the object with the highest rating is selected to be linked to the text.

Architecture of a geoparsing system

Architecture of a geoparsing system can be divided into two stages: data preparation and data presentation (Fig. 1). The data preparation stage is performed periodically, independently from user activities. For example, every hour the system connects to news feeds of external news sources and downloads their articles. These articles are then sent to the geoparser in order to identify potential geographic names which are then sent to geocoder. The goecoder queries the gazetteer database and retrieves place name identifiers for the analyzed text words that were found in the database. The geocoder then performs data ranking and links the words to the most highly ranked results. Finally, the texts with the linked geographic information are stored in a database until further retrieval for presentation purposes.

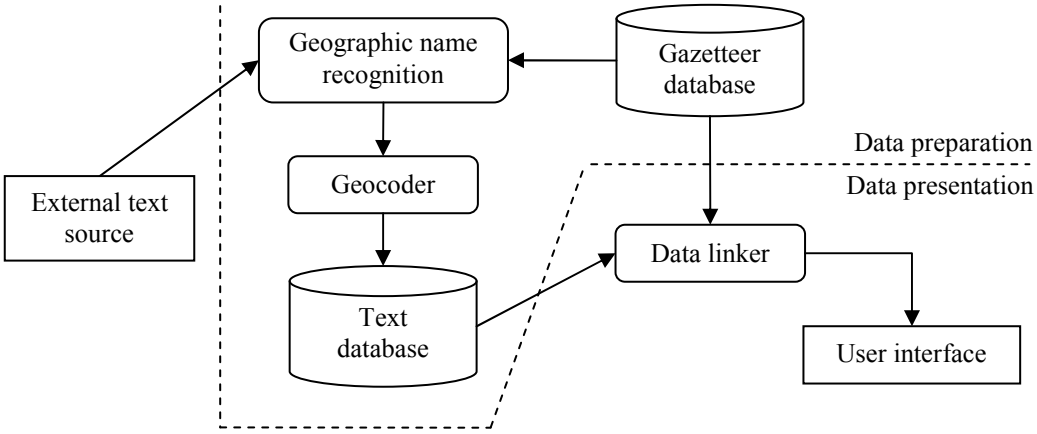


Fig. 1. **Overall architecture of textual data geoparsing system.**

In the data presentation stage, the texts from the database are retrieved, linking the additional information to the identifiers got in the previous stage. The text with attached additional information can then be shown in the user interface. The operations of the presentation stage are performed only on user request.

Developed prototype and its empirical evaluation

In this study, the developed geoparsing system prototype is oriented to the analysis of texts in Latvian language mentioning geographic places of Latvia. The used gazetteer consists of geographic names of Latvia only. The prototype supports both perfect and partial match database search methods, as well as potential geographic name filtering, word basic form generation, and ranking of found names. The prototype is implemented using Java programming language while as database storage open-source MySQL database management system was chosen.

For empirical evaluation of the developed prototype, a dataset of 20 real-world news articles (in Latvian and about Latvia) was created. The system was evaluated in two ways: recognition accuracy of individual geographic names and recognition accuracy for a whole article (in this case the whole text was to be linked to one geographic place to which the text refers the most).

Table 1 summarizes the results of the performed empirical evaluation. The table contains the following columns: the amount of correctly identified geographic names (in percents); the amount of correctly identified geographic names which exist in the database (in percents) – this is an important indicator to evaluate the recognition rate regardless of the completeness of the gazetteer database; the amount of incorrectly identified geographic names in relation to the total number of geographic names in the text; column with the answer “Yes” or “No” that shows whether the geoparser linked the whole text to the one correct geographic object. The results are calculated for both recognition methods – exact match and partial match, and using both, main names and alternative names from the gazetteer. In cases when exact match and partial match methods gave different results for the whole text linking, the last column contains two different answers: “(1)” when the result was obtained with exact match method and “(2)” when it was obtained with partial match method.

The results show that using the exact match method on average 80.14% of geographic names in the analyzed texts were found correctly. This percentage is considerably lower than 100% because geoparser made mistakes but most of the mistakes are due to the fact that the used gazetteer did not contain many of the places mentioned in the texts. If recognition rate is calculated only from those geographic names stored in the gazetteer, almost all of the names were identified correctly – an average of 98.33%. This suggests that the basic form generation algorithm is working very well – it has generated all basic forms correctly, but for future experiments and practical applications the gazetteer should be made more complete.

The amount of incorrectly identified geographic names (false positives) on average was 31.52% – the system incorrectly identified objects that actually were not mentioned in the text. This was because it identified places with the same name but different coordinates. For example, Gulbarga (in Gulbene district) and Gulbarga (in Daugavpils district) or Abava (in Ventspils district) and Abava (in Tukums district), etc. Note that in cases when the amount of such incorrectly identified names is larger that the total amount of geographic names in text, the calculated level may exceed 100%. Some news articles also mentioned more specific places, for example, one mentioned "Rīgas zooloģiskais dārzs". As the gazetteer database does not contain such name, the system only recognized city "Rīga" which in this case is only partly correct.

The percentage of correctly recognized geographic names using the partial match method is slightly higher (81.81%) because there were identified some geographic names due to their alternative names. Most recognition problems here are the same which were described in analysis for perfect match method. However, all the geographic names that were stored in the gazetteer database were successfully recognized.

Table 1

The results of empirical evaluation

Article	Exact match method			Partial match method			Whether the system has correctly linked the whole text
	Correctly identified geographic names, %	Correctly identified geo. names that exist in gazetteer, %	Incorrectly identified geographic names, %	Correctly identified geographic names, %	Correctly identified geo. names that exist in gazetteer, %	Incorrectly identified geographic names, %	
1	50%	100%	50%	50%	100%	100%	Yes
2	66,67%	100%	0%	66,67%	100%	100%	Yes(1),No(2)
3	100%	100%	8.33%	100%	100%	50%	Yes
4	50%	100%	0%	50%	100%	200%	Yes
5	100%	100%	200%	100%	100%	200%	Yes
6	100%	100%	0%	100%	100%	50%	Yes
7	50%	100%	0%	50%	100%	0%	Yes
8	100%	100%	33.33%	100%	100%	33.33%	Yes
9	100%	100%	0%	100%	100%	100%	Yes(partly)
10	83.33%	100%	0%	83.33%	100%	0%	-
11	100%	100%	0%	100%	100%	20%	Yes(partly)
12	60%	100%	40%	60%	100%	140%	No
13	66.67%	66.67%	33.33%	100%	100%	66.67%	Yes(2),No(1)
14	100%	100%	75%	100%	100%	125%	No
15	100%	100%	0%	100%	100%	50%	Yes
16	66.67%	100%	33.33%	66.67%	100%	33.3%	No(partly)
17	92.86%	100%	7.14%	92.86%	100%	14.29%	Yes(partly)
18	66.67%	100%	0%	66.67%	100%	33.33%	Yes
19	50%	100%	50%	50%	100%	100%	Yes
20	100%	100%	50%	100%	100%	25%	Yes
Average:	**80.14%**	**98.33%**	**31.52%**	**81.81%**	**100.00%**	**72.05%**	

This method shows much worse results in the incorrectly identified geographic names – the average percentage is 72.05%. The system has found almost as many incorrect objects as correct ones. This happens because many cities have lakes or other object names which at least partly match the name of the city (for example, a search for "Riga" also found "Riga Reservoir").

In the experiments, where the system was set to link the whole text to a one most appropriate geographic object, in 11 cases out of 20 the system perfectly identified the object using both methods. In two cases out of 20 one of the methods was wrong with the object choice. In three cases, the methods identified only partially correct answer but they can also be counted as correct answer because the task was to link the text to only one object while human would actually link the articles to two or more objects (and one of them the system identified correctly). In three cases, the system identified the object incorrectly. In one case, the system identified the village Abava while human would link this article to the river Abava (note that river names are not included in the used gazetteer). In one case, the system linked text to one particular city while a human could not link this article to any particular geographic object at all. To sum up all the results, the average percent of correct linking is equal to (11+1+3)/19 = 0.7895 or 78.95% but, taking in account that the database did not include information about rivers, the rate is (11+1+3)/18 = 0.8333 or 83.33%.

In order to quantify the differences in efficiency of the two word matching methods even more clearly, contingency tables were constructed (Table 2). From the tables, measures of relevance can be calculated, namely Precision (the fraction of retrieved instances that are relevant) and Recall (the fraction of relevant instances that are retrieved). For exact match method Precision is TP / (TP + FP) = 84.33% and Recall is TP / (TP + FN) = 80.46%. For partial match method Precision is 85.54% and Recall is 61.21%. As can be seen, Precision for this dataset for both methods is almost the same while for the partial match method Recall is considerably lower. The partial match method retrieves too many irrelevant results and therefore its usage is not recommended.

Table 2

Contingency table for recognition results of a) exact match method and b) partial match method

a)	**Geographic objects linked by system**	
Geographic objects to be linked	True Positive TP = 70	False Positive FP = 13
	False Negative FN = 17	True Negative TN = 66

b)	**Geographic objects linked by system**	
Geographic objects to be linked	True Positive TP = 71	False Positive FP = 12
	False Negative FN = 45	True Negative TN = 38

Conclusion

The empirical evaluation of the developed geoparsing system prototype showed that the average quality of its geographic name recognition varies around 75-80% if one takes into account the incompleteness of the gazetteer database used in experiments, while the percentage is close to 100% if one considers only the place names stored in the database. This suggests that for practical applications the developed geoparser could give sufficiently high degree of recognition but the quality of the used gazetteer database can significantly worsen this result. Increase in the amount of news articles used in experiments would increase the accuracy of the evaluation but the results would likely remain within the existing limits.

In order to improve the quality of place name recognition, it is necessary to improve the quality and completeness of the gazetteer database. It could also be helpful to add list of person name “stop words” as during the recognition process some of names and surnames mentioned in the articles were identified as geographic names. Also it would be helpful to enhance the gazetteer database with such large-scale objects as countries and municipalities and such small-scale objects as streets.

Future studies may include improvements of place ranking heuristics, improvements of gazetteer completeness, and comparisons of the system to other existing geoparsing systems.

Acknowledgements

The work has been supported by European Regional Development Fund project 2011/0008/2DP/2.1.1.1.0/10/APIA/VIAA/018 “Intelligent Agents, Modelling and Web Technologies Based Development of Distributed Insurance Software”.

References

Abascal-Mena, R., López-Ornelas, E., Zepeda, J.S., 2013. Geographic Information Retrieval and Visualization of Online Unstructured Documents. International Journal of Computer Information Systems and Industrial Management Applications, 5, pp. 89-97.

Caldwell, D., 2009. Geoparsing Maps the Future of Text Documents. Directions Magazine. Available at: www.directionsmag.com/articles/geoparsing-maps-the-future-of-text-documents/122487, 17.01.2012.

Goldberg, D.W., 2008. A Geocoding - Best Practices Guide. Springfield: NAACCR, pp. 261.

Keller, M., Freifeld, C., Brownstein, J., 2009. Automated vocabulary discovery for geo-parsing online epidemic Intelligence. Available at: www.biomedcentral.com/1471-2105/10/385, 03.02.2012.

Mikheev, A., Moens, M., Grover, C., 1999. Named Entity Recognition without Gazetteers. Proceedings of EACL ’99, Stroudsburg, pp. 1-8.

Nikolajevs, J., Jekabsons, G., 2011. Automated geocoding of textual data. Proceedings of Research innovations fundamentals 2011, Klapedos Universiteto, Klaipeda, pp. 12-16.

Elliptical Rule Extraction from a Trained Radial Basis Function Neural Network

Andrey Bondarenko, Arkady Borisov

DITF, Riga Technical University, Meza ¼, Riga, LV-1658, Latvia
Andrejs.bondarenko@gmail.com, arkadijs.borisovs@cs.rtu.lv

Abstract: *Currently knowledge management and discovery plays crucial role in different industries. Apart from that in many cases decisions made by companies in mission critical areas (like nuclear power industry, medicine and finance, to name a few) should be verified by domain expert and/or explained. This is needed to fulfill regulatory requirements, to verify correctness and / or discover new knowledge. Artificial neural networks are well known models that can be used to solve classification problems. Unfortunately they are "black box" models for end users, because classification process is fully hidden from the observer and cannot be easily mapped into formalized meaningful form that can be understood by the domain expert. There are multiple knowledge representations, and the one chosen to work with in this study is elliptical rules. A set of such rules can be used to determine the class of an input data point by the determination of ellipsoid that covers provided point. Current paper addresses the problem of elliptical rules (ER) extraction from trained artificial radial basis function neural network (RBFNN). This study uses RBFNN with tunable nodes - such networks are built using orthogonal forward feature selection algorithm and they usually have much less neurons (in comparison to RBFNNs containing neurons with fixed radii) while having comparable or even superior accuracy. The article poses non-convex optimization problem of finding multiple ellipsoids of largest volume inscribed into RBFNN decision boundary. Non-convexity arises due to complex nature of RBFNN decision boundary which serves as constraint. Further we describe an algorithm for extraction of ER from trained RBFNN. The provided experimental results show that few of the extracted elliptical rules have comparable and in some cases higher classification accuracy than original RBFNN. Although curse of dimensionality is applicable to the provided algorithm, experiments have shown that it can be readily used for relatively low-dimensional problems. Finally are described the possible future research directions.*

Keywords: radial basis function networks, knowledge acquisition, optimization.

Introduction

There are many problems for which artificial neural networks are still a preferable solution. They can be trained on data with outliers and are natural choice for multi classification problems, although the way in which classification is performed is a black-box algorithm for the end user. Due to specific requirements of different industries for clear and formal decision process description for validation and knowledge gathering purposes, knowledge extraction from trained neural network is a topical problem. In this paper we propose an algorithm for the extraction of elliptical rules from trained artificial radial basis function neural network with tunable nodes (Chen et al., 2009). Radial basis function neural networks are local in their nature - meaning each neuron is responsible for classification decision in specific region based on defined neuron radius and location parameters (and weight). Having that, it is possible to define optimization problem that would allow us to cover classification region with ellipsoids - hence giving us a formalized view of how classification is made. The structure of the paper is as follows: Chapter 2 provides background on neural networks, Chapter 3 poses optimization problem and algorithm, Chapter 4 shows experimental results and Chapter 5 concludes.

RBF Neural Network With Tunable Nodes

Radial basis function artificial neural network (Moody et al., 1989; Poggio et al., 1990) can be represented as follows:

$$(x) = \sum_{i}^{N} a_i p(\|x - c_i\|) \quad \textbf{(1)}$$

where a_i is the *i*-th neuron weight, c_i is the *i*-th neuron center, *N* is neurons count. The norm is usually taken to be Euclidean distance and the basis function *p* is taken to be Gaussian:

$$p(\|x - c_i\|) = \exp(-\beta\|x - c_i\|^2) \quad \textbf{(2)}$$

There are multiple strategies for training this kind of network. The network can have neurons with fixed radii, as this simplifies learning procedure. On the other hand, having neurons with different radii reduces the amount of neurons required to get necessary classification accuracy. But having neurons with varying radii requires a

specialized learning approach. We used the method described in (Chen et al., 2009). The proposed method allows us to build RBF networks with a small amount of neurons while preserving high accuracy.

Rule Extraction

Optimization problem

We can treat elliptical rule extraction from trained RBF neural network can as an optimization problem of finding ellipsoids of maximum volume inscribed into the space area(-s) defined by RFBNN decision boundary. It is possible to choose maximization criteria other than volume- like amount of points covered by newly shaped ellipsoid, but in this case we will not be able to use gradient-based solvers.

Let's denote ellipsoid as

$$\varepsilon = \left\{ Bu + d \mid \|u\|_2 \right\} \tag{3}$$

a unit ball under affine transformations. As described in (Boyd and Vandenberghe, 2004) we say that *B* is n-length vector containing positive elements, so ellipsoid volume is proportional to *detB*. We can write down the following optimization problem:

$$\begin{aligned} &\max \log(\det B)) \\ &s.t. RBNN \supseteq \varepsilon \end{aligned} \tag{4}$$

This means – we are looking for ellipsoid of maximum volume fully inscribed into RBFNN defined decision boundary. Apart from that we have explicit constraint that ellipsoid should have non-negative radii – elements of vector B. The described problem allows us to find first ellipsoid inscribed into the RBFNN decision boundary. We should note that due to RBFNN nature, constraints of our problem are non-convex, thus the found ellipsoid can be a local solution. In most cases it will be insufficient to represent RBFNN with a single ellipsoid, thus we need to search for additional ellipsoids.

For that we need to apply the iterative approach. In contrast to recursive volume subdivision applied in (Nunez et al., 2002), we have chosen another strategy. Although recursive subdivision is also a feasible approach as it does not require objective function modification, on the other hand, it generates larger count of ellipsoids. Instead of space subdivision for searchable regions on each recursive iteration, we are just looking for inscribed ellipsoid with maximum volume not covered by already found ellipsoids, meaning the larger fraction of ellipsoid lies outside the region already covered by existing ellipsoids, the better it fits our needs:

$$\begin{aligned} &\max(\varepsilon_{vol} - E_{vol}) - P \\ &s.t. RBNN \supseteq \varepsilon \end{aligned} \tag{5}$$

here ε_{vol} denotes the volume of found ellipsoid and E_{vol} is the volume of already existing ellipsoids and *P* is penalty term. Modification in terms of volume calculation introduced large plateau areas with objective function equal to zero. To allow faster convergence of optimization procedure, penalty term *P* calculates minimal distance between candidate ellipsoid center and border of set formed by intersection of all previously found ellipsoids. Thus the 'further' the center of candidate ellipsoid contained within other found ellipsoids is, the larger penalty term *P* will be.

Such modifications allow us to ensure that, on each optimization iteration convergence will run faster and will allow us to find a new ellipsoid, which will cover most of not yet covered volume. Of course, there is no guarantee that the found ellipsoid will be a global solution.

Algorithm

Here we will describe an algorithm for elliptical rule extraction from a trained RBFNN. It is necessary to mention that extracted ellipsoids will be oriented parallel to coordinate axes, so original unit ball is deformed only by stretching, but not by rotation. On the very first iteration we need to find an inscribed ellipsoid of maximum volume that fits into the RBFNN decision boundary. In this case, objective function is defined as:

$$\log(prod(B)) \tag{6}$$

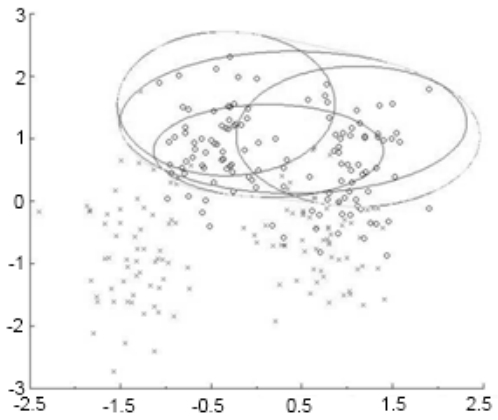


Fig. 1. **Decision boundary of RBFNN with 2 neurons and 2 extracted ellipsoids.**

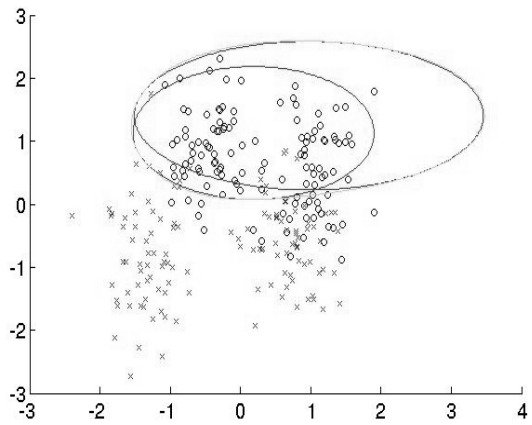


Fig. 2. **Decision boundary of RBFNN with 6 neurons and 4 extracted ellipsoids.**

The algorithm is listed below and has these inputs: *Data* - training input vectors, *C, R, w* - are parameters depicting RBFNN from which will be decomposed into the elliptical rules. The output of the algorithm is *Ellipses* set which contains ellipsoids. In the algorithm description *cRBF* – is handle for constraint function RBF which accepts *C*, *R* and *w* which are neurons centers, radii and weights respectively; *ub, lb* and *x0* – are the upper bound, lower bound and initial starting point for optimization. o*bjVol* – is objective function which accepts ellipsoid vector containing ellipsoid radii and ellipsoid center vector. cR*bfMod* and *objVolMod* are constraints and objective function modified versions. *U* – are points covered by RBFNN, but not covered by the supplied set of ellipsoids. *n* – is the amount of points (from *U*) that are not covered by newly found ellipsoid. If *n* is 0, then algorithm ads newly found ellipsoid and returns the result. This version of the algorithm is different than that provided in (Bondarenko and Borisov, 2012). Although it produces the same results in a two-dimensional case, it is capable of generating only a single inscribed ellipsoid. In *haberman* data set we stumbled on cases when a single ellipsoid was enough to cover data all points that are located inside RBF decision boundary. Thus we slightly modified the algorithm allowing it to terminate after acquisition of the first ellipsoid. Another point not depicted in the algorithm below is that in one of the two examined cases the algorithm produced four ellipsoids (this was set by maximum allowed ellipsoids count variable), but their manual clean-up revealed that only two of them are sufficient for covering all, but two points covered by RBFNN being decomposed. Such clean-up cannot be easily incorporated into the algorithm itself as each found ellipsoid directs global search into new not covered volumes, thus even if ellipsoid is not covering any new not yet covered points it should be preserved. In case when genetic algorithm or multi-start global search are used this should not be a problem.

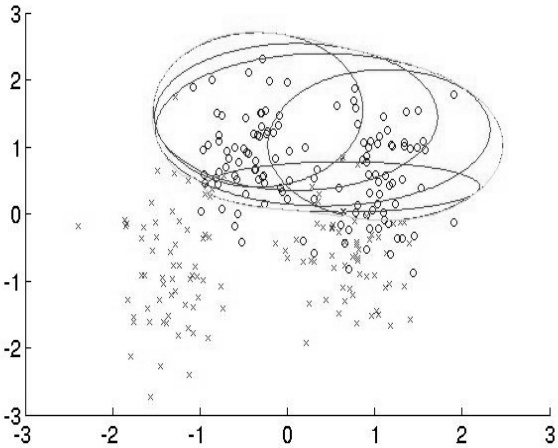


Fig. 3. **Decision boundary of RBFNN with 6 neurons and 5 extracted ellipsoids.**

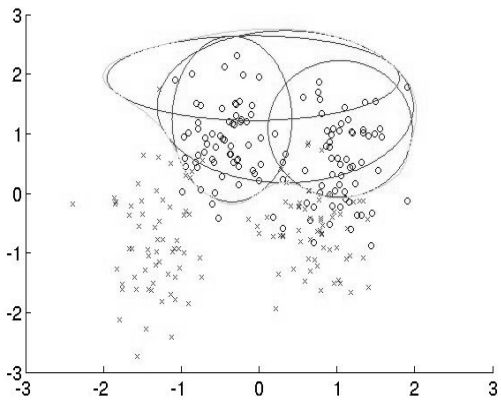


Fig. 4. **Decision boundary of RBFNN with 7 neurons and 4 extracted ellipsoids.**

```
IN:  ( maxEllipsoidsCount, Data, C, R, w )
OUT: ( Ellipsoids )

cRbf = @(x) constraintRbf(x, C, R, w);
objVol = @(x) objVolume(x);
e1 = solve(cRbf, objVol, ub, lb, x0);
Ellipsoids = e1;

U = uncoveredPoints(Data, Ellipsoids, C, R, w);
if (count(U) == 0)
   return e1;
end

i = 2;
while i < maxEllipsoidsCount
       i++;
       U = uncoveredPoints(Data, Ellipsoids, C, R, w);
       cRbfMod = @(x) constraintRbfModified(x, C, R, w);
       objVolMod = @(x) objVolumeModified(x);
       e = solve(cRbfMod, objVolMod, ub, lb, x0);
       n = size(uncoveredPoints(U, e, C, R, w), 1);
  Ellipsoids = Ellipsoids + e;
  if (n == 0)
     break;
  end
end
```

Table 1

RBFNN Accuracy, Extracted Ellipsoids Accuracy and Count

# of Neurons in RBFNN	RBNN Train Accuracy$_{(std.dev.)}$	RBNN Test Accuracy$_{(std.dev.)}$	Ellipsoids Train Accuracy$_{(std.dev.)}$	Ellipsoids Test Accuracy$_{(std.dev.)}$	Ellipsoids count $mean_{min}^{max}$
2 neurons	0.852	0.911	$0.84_{0.000}$	$0.887_{0.000}$	2_2^2
6 neurons	0.868	0.905	$0.868_{0.002}$	$0.9032_{0.005}$	4.4_5^4
7 neurons	0.876	0.905	$0.876_{0.000}$	$0.9031_{0.002}$	5.1_7^4
9 neurons	0.868	0.905	$0.8728_{0.005}$	$0.9039_{0.001}$	6.8_7^5

Experiments

We have created an algorithm supporting two and three dimensional input space. Furthermore, our final goal was to show that extracted rules are approximating RBFNN as close as possible. We set up experiments on synthetic two-dimensional Ripley data set, which can be found at (Frank and Asuncion, 2012) as well as haberman survival data set. Both of them have two classes. We run RBFNN construction algorithm described in (Chen et al., 2009) to construct several neural networks containing different amounts of neurons. Furthermore we observed only closed RBFNN defined classification boundaries which can be seen in figures. Looking at the algorithm one can notice *maxEllipsoidsCount* variable. We've initialized it with the number of neurons in the subject RBFNN, the only exception was a network with 9 neurons, for which maximum ellipsoids count to be extracted was set to seven as well as three dimensional data case where we set that value to three ellipsoids.
As it was already noted, the algorithm was not executed on open (not bounded areas meaning decision boundary lies outside the lower and upper bounds) decision areas, and RBFNN decision boundaries consisting of several separate space volumes (like two neurons forming two separate positive decision areas). Decision boundaries and extracted ellipses can be observed in Fig. 1-8 for two dimensional case and Fig. 9-10 for three-dimensional case. Experimental results can be observed in Table 1. One can notice that testing accuracies are higher than the training ones; this is due to the nature of training/testing data being used. We have not collected accuracies for three-dimensional cases apart the fact that ellipsoids built for the case depicted in Fig.9 left 3 uncovered points, while 2 ellipsoids depicted in Fig. 10 covered all data points that lied inside RBFNN decision boundary.

Overall computation time was partially an issue for us, finding the very first ellipsoid turned out to be rather quick operation while searching for subsequent ellipsoids was more CPU intensive task due to the amount of

computations needed to calculate modified objective function. Another point to mention is algorithms used in volume intersection and ellipsoid containment with RBFNN boundary calculations. In case of 2 dimensions we utilized mapping / clipping algorithms that allow us to acquire figure which corresponds to area belonging to candidate ellipse laying outside of already found ellipses. In case of three dimensions, we utilized an approach similar to the one described in (Persson and Strang, 2004). We defined distance-based volumes which were translated into three-dimensional meshes. This implied selection of appropriate mesh grid step for construction of isosurface for which volume can be calculated with appropriate precision. For checking whether ellipsoid is fully included within RBFNN decision boundary we created a set of points on its surface and checked each of points to be belonging to the required bounded volume. Although one can use any of the global optimization approaches, we believe GA is not a good option here, while search involving local solvers with different starting points proved to be the best in terms of computation time. Overall accuracy of the extracted rules – ellipses in 3 out of 4 cases lies within 1% which is a good result, taking into account the amount of rules being extracted.

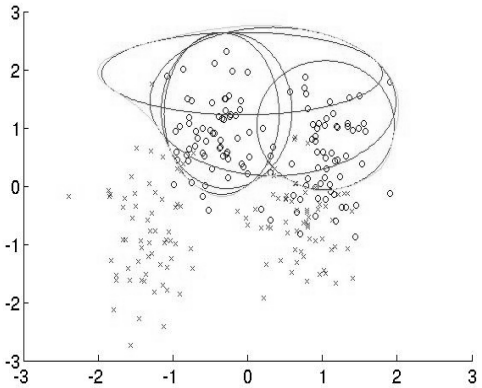


Fig. 5. **Decision boundary of RBFNN with 7 neurons and 5 extracted ellipsoids.**

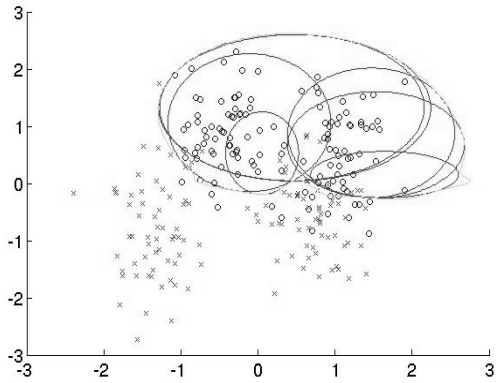


Fig. 6. **Decision boundary of RBFNN with 9 neurons and 7 extracted ellipsoids.**

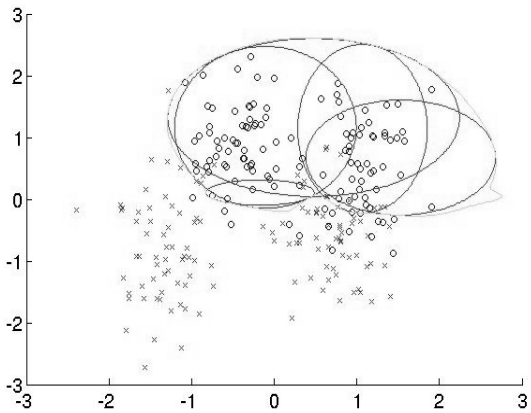


Fig. 7. **Decision boundary of RBFNN with 9 neurons and 5 extracted ellipsoids.**

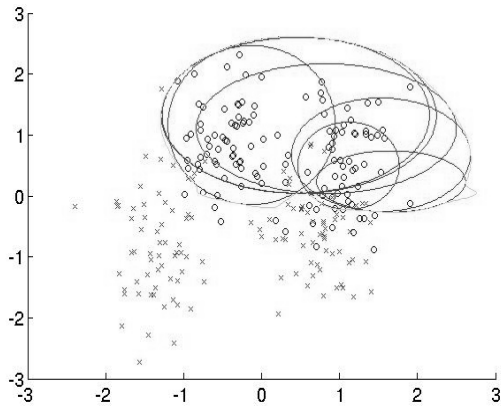


Fig. 8. **Decision boundary of RBFNN with 9 neurons and 7 extracted ellipsoids.**

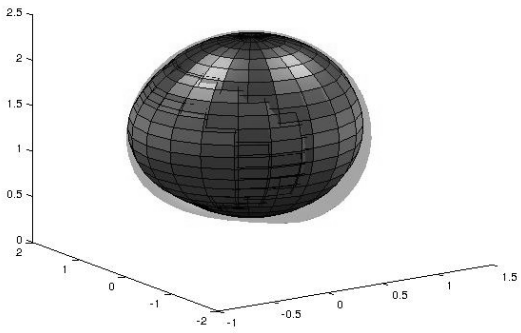


Fig. 9. **Three dimensional case. Decision boundary of RBFNN with 2 neurons and 3 extracted ellipsoids.**

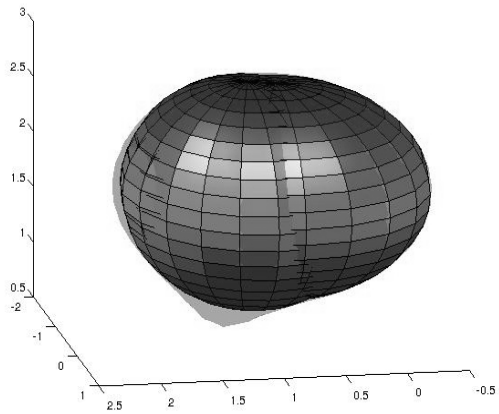


Fig. 10. **Three-dimensional case. Decision boundary of RBFNN with 2 neurons and 2 extracted ellipsoids.**

We believe it is possible to lower execution times by introducing heuristics for initial point selection. Thus basically we can deal with the acquisition of local solutions which can be better approach than the utilization of multistart optimization approach. In our algorithm we have used an ellipse fully residing inside RBFNN decision boundary without applying additional constraints; though constraints relaxation can potentially lower the count of extracted rules.

Conclusion

We have investigated the possibility of elliptical rule extraction from radial basis function neural networks. We developed an algorithm which was successfully applied to two and three-dimensional input data and radial basis function neural network trained on that data. The results observed indicate that the proposed algorithm can be successfully applied to low-dimensional problems. The experiments have shown that computation time can be a problem especially in case of larger RBF neural networks.

Apart from that feasible research direction is RBF structure exploiting to speed up constraints calculation, along with that algorithm is not tested on RBFNN decision boundary which covers open sets and several isolated space regions. Here by saying an open set we mean that classification boundary partially lies behind the upper and lower domain boundaries. This can be solved by extending boundaries further away or even eliminating them, although the effect of that needs to be tested and clearly finding appropriate solution depends on the optimization algorithm used. As it was noted in the Chapter Rule Extraction, one can utilize recursive space subdivision for subsequent searches although this can imply some limitations on fitness of found ellipsoids.

Overall amount of found ellipsoids along with the demonstrated accuracy shows that the proposed approach is feasible, especially for small radial basis function neural networks with low-to-moderately sized RBF layer.

References

Moody, J.and Darken, C.J., 1989. Fast learning in networks of locally tuned processing units, *Neural Computation*, 1, pp.281-294.

Poggio, T.and Girosi, F., 1990. Networks for approximation and learning, Proc. IEEE 78(9), pp.1484-1487.

Chen, S.X., Hong, Luk, B.L. and Harris, C.J., 2009. Construction of tunable radial basis function networks using orthogonal forward selection, *IEEE Trans. Systems, Man, and Cybernetics, Part B*, Vol.39, No.2, pp.457-466.

Bondarenko, A. and Jumutc, V., 2011. Extraction of Interpretable Rules from Piecewise-Linear Approximation of a Nonlinear Classifier Using Clustering-Based Decomposition, Cambridge, AIKED.

Nunez, H., Angulo, C., Catala, A., 2002. Rule extraction from support vector machines, Proceedings of the 10th European Symposium on Artifical Neural Networks, pp.335-343, Bruges, Belgium, April24-26.

Boyd, S., Vandenberghe, L., 2004. *Convex Optimization,* Cambridge University Press.

Bondarenko, A. and Borisov, A., 2012. Elliptical Rules Extraction from Trained Radial Basis Function Neural Network, Scientific Journal of Riga Technical University.

Frank, A.and Asuncion, A., 2012. *UCI Machine Learning Repository,* Available at: http://archive.ics.uci.edu/ml, 29, 2012. School of Information and Computer Science, Irvine, University of California.

Persson, P.O. and Strang., G.,2004. *A Simple Mesh Generator in Matlab*, SIAM Review, June, Volume 46(2), pp. 329-345.

Clustering algorithm specifics in class decomposition

Inese Polaka

Institute of Information Technology, Riga Technical University, Meza street 1/3, Riga, LV-1048, Latvia
Inese.Polaka@rtu.lv

Abstract: *The task of the presented study is to find different disease phenotypes of cancer (breast cancer, carcinoma, gastric cancer, melanoma, prostate cancer) and gastrointestinal inflammatory disease using clustering algorithms. The article analyzes the performance of two different approaches to clustering data for class decomposition. One of them is using agglomerative hierarchical clustering and analyzing the obtained dendrogram to determine the number of disease subtypes; another is using k-means algorithm and determining the number of disease subtypes by analyzing the cluster compactness after several runs (using different numbers of clusters/cluster centers). After clustering is done, the clusters are analyzed to assess their specifics and the potential of clusters being different phenotypes or disease subtypes. The initial analysis of clustering results consisted of analyzing records belonging to clusters, cluster sizes and specifics. The secondary analysis of clustering was cluster quality evaluation that was done using classification algorithms (C4.5, Random Forest and SVM). The hypothesis is that well formed clusters would create disease subtypes that would be easily split using classification algorithms. The main results of the study show that the secondary analysis of the clusters is very similar for both clustering approaches and increases the classification results compared to the results of initial full data classification. The results also point to the sensitivity of k-means algorithm to noise and outliers because the initial analysis showed that the clusters formed a main group of records and several clusters of very few records. Although hierarchical clustering asks for expert opinion in cluster number determination, it also showed that it formed several large clusters that could point to phenotypical subtypes of the diseases.*

Keywords: bioinformatics, class decomposition, clustering.

Introduction

Finding the correct and most effective algorithms is a significant task of bioinformatics. This article takes on the task of biomedical diagnosis by implementing machine learning algorithms to solve a classification task discriminating between two classes – affected patient or a healthy individual. To improve the performance of the classification algorithms, the classification process involves using inner class density structures which are thought to represent disease subtypes – different phenotypes of the same disease.
In this article it is a clustering task to decompose classes into subclasses revealing disease subtypes. And this study analyzes the impact of clustering algorithms and their specifics on the task of class decomposition.

Methods

The unknown phenotypes of the cancer diseases are defined as high density areas in the attribute space that can be found by clustering algorithms. This article examines two different approaches to clustering: partitioning (k-means algorithm) and divisive (hierarchical agglomerative clustering algorithm). The algorithms are described in the following subsections. Also briefly described in this section are classification algorithms as well as the used classification result validation technique and result interpretation.
The data used in the study are provided by Latvian Biomedical Study and Research Center (antibody phage display data for breast cancer (BrCa), gastric cancer (GaCa), gastrointestinal inflammatory disease (GIS), melanoma (Mel) and prostate cancer (PrCa) patients and healthy donors) and acquired from the Internet (gene expression data of breast cancer (BC1, BC2), carcinoma (Carc) and prostate cancer (Pr) patients and healthy donors).

Hierarchical agglomerative clustering

Hierarchical clustering algorithms either merge objects (and clusters) until all objects belong to one cluster (agglomerative or bottom-up hierarchical clustering) or split clusters until each cluster holds exactly one object (divisive or top-down hierarchical clustering). The process can be effectively pictured as a dendrogram showing all splits or merges as shown in Fig. 1.

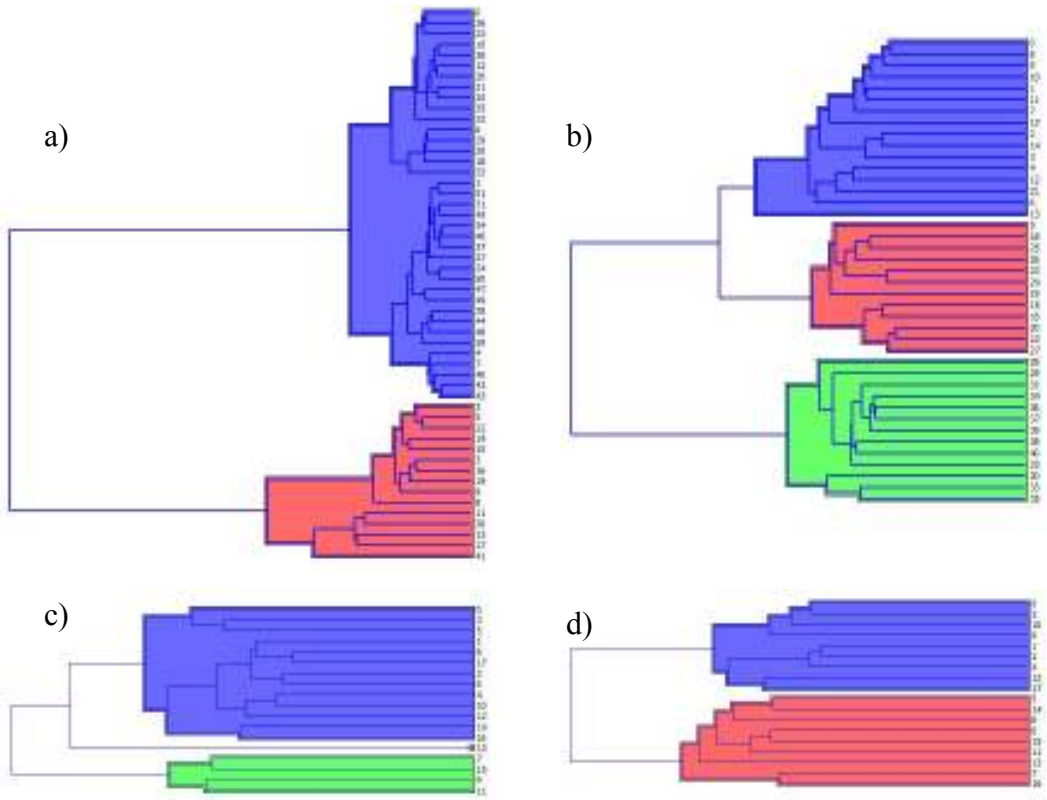


Fig. 1. **Dendrograms of hierarchical agglomerative clustering:** a) Prostate cancer antibody display data; b) Breast cancer (2) gene microarray data; c) Breast cancer (1) gene microarray data; d) Carcinoma gene microarray data.

The merging and splitting is applied to the objects (clusters) that are the closest to each other. There can be different distance measures and also between-cluster distance can be calculated in different ways. First let us take a look at the most popular distance measures between objects:

- Euclidean distance:

$$d_E(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (1)$$

where q and p are two objects.

- Manhattan distance:

$$d_M(q,p) = \sum_{i=1}^{n}|q_i - p_i| \quad (2)$$

- Correlation based distances (Pearson's correlation):

$$d_P(q,p) = \frac{\sum_{i=1}^{n}(q_i-q)(p_i-p)}{\sqrt{\sum_{i=1}^{n}(q_i-q)^2}\sqrt{\sum_{i=1}^{n}(p_i-p)^2}} \quad (3)$$

Using one of these metrics the distance is calculated between two objects but there is another way of measuring distance between a cluster and an object or between two clusters, taking into account records of the cluster and optimizing theobjective function (Manning et al., 2008):

- Single linkage takes into account only vectors of the two most similar records of a cluster (the closest points of the clusters in the attribute space);
- Complete linkage takes into account only the two most dissimilar objects (the furthest points of the clusters in the attribute space);
- Average linkage takes average distance between objects from the first cluster and objects from the second cluster as the distance between clusters (Koliadin, 2012):

$$D_A(X,Y) = \frac{1}{N_X \cdot N_Y}\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} d(x_i,y_j) \quad (4)$$

where X and Y are clusters,
N_X and N_Y are numbers of objects in clusters X and Y,
$d(x,y)$ is the distance between objects.

- Ward's minimum variance method chooses for merging the two clusters whose merge will result in the minimum within-cluster variance.

The algorithm used in this study uses Euclidean distance and Ward's between-cluster distance because the latter points to possible subclasses by using a similar metric as in classification algorithm and this method usually uses Euclidean distance for regular data sets.
The number of clusters is usually determined by making a cut in the dendrogram (Fig. 1) by cutting the longest branches, which means choosing the number of clusters so that the distance between the merges of these clusters in the dendrogram is the longest.

K-means clustering

K-means algorithm partitions a data set into *k* clusters where the number *k* is predefined or found using trial clustering. The partitioning process is carried out so that the similarity between two records in a cluster is higher than similarity between two records that belong to different clusters. The clusters are represented by their centroids (center points that are calculated as the mean value of the objects in the clusters).
The clustering process is iterative. In the initial step the *k* centroids are chosen from the record set. Then iteratively each object is assigned (reassigned) to the closest cluster based on the cluster centers and, when all objects belong to clusters, the cluster centers are calculated as mean values of the objects belonging to clusters as presented in Equation 1 (Han and Kamber, 2006).

$$\vec{\mu}(\omega) = \frac{1}{|\omega|}\sum_{\vec{x}\in\omega}\vec{x}, \tag{5}$$

where ω – cluster identifier,
$\vec{x}$ – value vector of an object belonging to cluster ω.

The clusters are evaluated using the residual sum of squares (RSS) as a distance measure. For each cluster *k* it is calculated as shown in Equation 2.

$$RSS_k = \sum_{\vec{x}\in\omega_k}|\vec{x} - \vec{\mu}(\omega_k)|^2 \tag{6}$$

The process is repeated until a stopping criterion is met. The stopping criterion can be one of the following (Manning et al., 2008):

- A fixed number of iterations;
- Assignment of objects to clusters does not change;
- Centroids do not change between iterations;
- The residual sum of squares (RSS) decreases below threshold or the decrease is smaller than a threshold.

Whereas the number of clusters *k* has to be given a priori, in the experiments of this study it is chosen by applying the algorithm to the data several times and changing the number of clusters. Then for each number of clusters the mean distance between clusters is calculated. The number of clusters *k* is chosen so that the distance between clusters is the lagest meaning clusters being less similar. The values and the best choices of *k* are shown in Table 1.

Table 1

Distances between clusters in k-means clustering

k	BC1	BC2	Carc	Pr	BrCa	GaCa	GIS	Mel	PrCa
2	842.0	1449.9	601.1	2567.4	195.3	1104.0	1029.3	1032.9	487.2
3	831.5	1142.3	489.6	1258.8	149.5	908.9	743.3	1167.5	363.8
4	826.0	1139.6	520.8	1051.9	156.3	870.5	711.0	973.9	364.9
5	790.1	1205.2	475.8	1058.1	154.2	955.2	706.1	945.1	364.1
6	689.2	1150.9	470.7	1054.3	152.5	768.8	640.6	925.2	348.1
7	688.4	1175.0	482.0	990.1	153.5	781.9	661.4	909.6	344.5
8	702.8	1129.6	471.9	946.5	147.7	738.6	661.7	912.4	358.7
9	693.9	1124.3	466.3	1038.0	150.0	798.6	648.5	869.0	344.2
10	700.5	1109.4	486.9	952.6	144.0	767.6	608.4	883.7	337.0

Classification algorithms

Classification algorithms used in this study are the ones that are the most popular and recognized as very well interpretable and the most effective in other studies (Polaka et al., 2010; Lee, 2005; Lu and Han, 2003; Dudoit et al., 2002) – algorithm C4.5, Random Forests (RF) and Support Vector Machines (SVM).

Algorithm C4.5 usually uses Information gain or Gain ratio as the criteria to choose the attribute for each split. Information gain is the change in entropy of information if the state of information is changed. Let *C* be the class attribute with values $\{c_1, c_2, \ldots, c_n\}$ and *A* attribute with values $\{a_1, a_2, \ldots, a_k\}$, *H(C)* be the entropy of the class attribute, and *H(C|A)* conditional entropy that shows entropy of *C* if state of attribute *A* is known, Information gain is:

$$I(C,A) = H(C) - H(C|A) \tag{7}$$

The entropy of attribute *C* is:

$$H(C) = -\sum_{i=1}^{n} P(C = c_i) log_2 (P(C = c_i)) \tag{8}$$

where $P(C=c_n)$ is the relative frequency of class value c_n. And the conditional entropy is:

$$H(C|A) = -\sum_{j=1}^{k} P(A = a_j) H(C|A = a_j). \tag{9}$$

Information gain favors attribute with higher number of values. To avoid that, gain ratio can be used. This criterion penalizes a large number of attribute values by dividing Information gain with entropy of the attribute itself:

$$IG(C,A) = \frac{I(C,A)}{H(A)} \tag{10}$$

where the entropy of attribute A is calculated as follows:

$$H(A) = -\sum_{j=1}^{k} P(A = a_j) log_2 (P(A = a_j)). \tag{11}$$

Random forests use a large number of unpruned decision trees, which are created by randomizing the split at each node of the decision tree. The number of attributes used to determine the decision at a node of the tree is predefined and is less than the original number of attributes. The attributes are chosen randomly and the best split among those attributes is chosen. The classification of a new sample is performed using majority vote (Rokach, 2008).

Support vector machine (SVM) builds a function of relevant features by assigning weights to them (irrelevant features are assigned weight 0) based on relevant instances (support vectors). The function is a hyperplane in the instance space that separates different classes with a maximum margin (distance from the hyperplane to the nearest instances). SVMs have various types and enhancements; this study employs an enhancement called *Sequential Minimal Optimization* (SMO) introduced by Platt that is used for training support vector classifiers (Platt, 1998). It was also improved by Keerthi and Shevade (Keerthi, 2001). This approach breaks training process into smaller, two-dimensional problems and reduces resource consumption comparing to large matrix computation needed for the classic SVM training. SVMs also use kernels to transform feature spaces where they search for hyperplanes. In this study the Polynomial kernel was used to represent dot products.
While SVMs only work with binary classes, the multi-class problem is solved using pairwise classification 1-vs-1 (*pairwise coupling method*) proposed by Hastie and Tibshirani (Hastie and Tibshirani, 1998).

Result validation and interpretation

The classification results are validated using 10-fold cross-validation. This means that initially each data set is divided into ten parts and then alternately each of the parts is left out for testing while nine other parts are used for training. This enables results validation as if there were ten data sets available while not asking for any supplemental records – and each record in this study means an expensive antibody display or gene microarray.
The results are interpretd for the initial classes instead of the found subclasses to go on with the scientifically proven classes, which are also most relevant – either a patient has the disease or not. The objects that are assigned to the positive subclasses are summed up to make up the positive class and the negative class is kept as distinguished in the subclass classification, which is the same as in the initial data set (it is not decomposed).

Results and discussion

First all data sets were split according to classes for class decomposition because only the positive class was decomposed following the hypothesis that class decomposition finds disease subtypes with different phenotypes. Next these subsets of positive instances were decomposed revealing the inner class density structures by applying clustering algorithms. The cluster structures acquired by k-means algorithm are presented in Fig. 2, while cluster structures of hierarchical agglomerative clustering are shown in Fig. 3.
In Fig. 2 C1, C2 and C3 are the labels of the clusters that the records were divided into. The same notation is used in Fig. 3 that shows how the records were divided into clusters using hierarchical agglomerative clustering algorithm.
As it can be seen from Fig. 2, there is one dominating cluster in all data sets while the other clusters hold ~10% of data. This is typical for outliers – the main cluster is formed in the regular data area while other clusters are moved to the extreme values and hold only some outliers. This is also typical for k-means algorithm – it tries to move cluster centers further apart from each other and this way other cluster centers were moved to the marginal values.

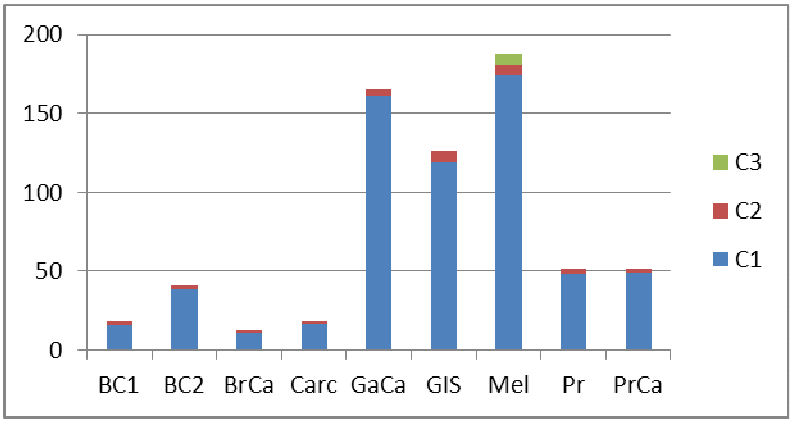


Fig. 2. **Distribution of records among clusters when k-means clustering is applied**.

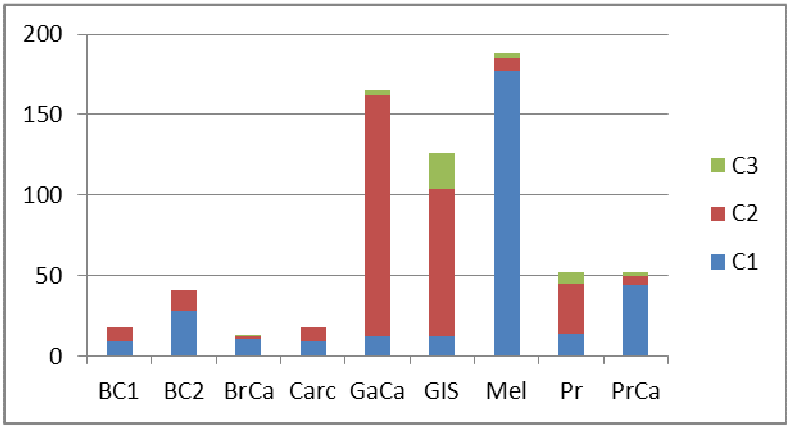


Fig. 3. **Distribution of records among clusters when hierarchical agglomerative clustering is applied**.

Fig. 3 shows that there is also a dominating cluster in the most data sets but other clusters are more significant because instead of holding a couple of outfitters, they consist of more records that could point to more useful high density areas in the data. Although there is also a tendency in hierarchical clustering to make further records a different cluster, it is not as strong because it points to higher density areas.

After revealing the inner class structures, the data subsets of different classes were merged with the initial negative subsets. These acquired sets were then used for classification. All three classification algorithms were used for all data sets to compare classification algorithm efficacy. Whereas assessment of the class decomposition impact was carried out by comparing results of different clustering algorithms and comparing each of them to the benchmark classification results, which were obtained by classifying the initial data sets as it is classically done in classification tasks.

The benchmark accuracies that were obtained by classifying initial (unprocessed) data sets and validating them using 10-fold cross-validation are shown in Table 2.

Table 2

Benchmark classification accuracies.

	C4.5	**RF**	**SVM**
BC1	64,29	57,14	59,52
BC2	64,58	67,71	79,17
BrCa	57,69	84,62	88,45
Carc	91,67	91,67	97,22
GaCa	59,38	58,75	66,88
GIS	49,64	49,64	58,57
Pr	85,29	79,41	91,18
PrCa	83,50	85,50	88,50

It is obvious that in seven cases out of eight the best accuracies are achieved by SVM classifier whereas results of both tree-based classification algorithms are very similar in most cases.

The classification accuracies obtained using data sets with class decomposition that was carried out using k-means algorithm are presented in Table 3. The results show that in 17 cases out of 24 the classification accuracy has increased (underlined results) even reaching perfect classification (100%).

Table 3

Classification accuracies after class decomposition (k-means).

	C4.5	**RF**	**SVM**
BC1	66,67	59,52	66,67
BC2	69,79	100,0	57,29
BrCa	80,00	91,67	93,33
Carc	86,11	97,22	97,22
GaCa	57,19	58,75	65,00
GIS	52,50	83,93	57,86
Pr	82,35	80,39	93,14
PrCa	72,50	96,00	91,50

The classification results for the data set where classes were decomposed using hierarchical clustering are shown in Table 4. It is also obvious that the accuracies in the most (16 out of 24) experiments have also increased (underlined results). In the most data sets the best result is obtained using the decomposed data sets that used k-means clustering but, as it was shown in the cluster breakdown previously, most of the increase is due to the identified outliers, which is not as relevant for the task.

Table 4

Classification accuracies after class decomposition (hierarchical agglomerative).

	C4.5	**RF**	**SVM**
BC1	69,05	66,67	64,29
BC2	66,67	70,83	75
BrCa	73,08	57,69	88,46
Carc	91,67	100	97,22
GaCa	61,88	55,94	67,19
GIS	63,57	55,87	63,93
Pr	72,57	75,49	90,2
PrCa	75	81,5	91

Conclusion

The results of this study show that there are differences between classification results for data sets decomposed by different clustering algorithms. While there is essential improvement in classification accuracies because of class decomposition, the most effective clustering algorithm is not obvious. Although application of k-means clustering algorithm leads to more improvements in classification results, there is a strong suspicion that it is due to outlier detection. The cluster structures showed that there is one main cluster in k-means clustering and other clusters hold less than 10% of the results, whereas hierarchical clustering showed more significant clusters pointing to high density areas in attribute space that could be the disease subtypes that were the main goal of class decomposition.

Acknowledgements

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».We thank the organizers of conference

References

Dudoit, S., Fridlyand, J., Speed, T. P. 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97- 457, pp. 77–87.

Han, J. and Kamber, M. 2006. *Data Mining: Concepts and Techniques, 2nd ed.*, The Morgan Kaufmann Series in Data Management Systems, Jim Gray (Ed.), Morgan Kaufmann Publishers, San Francisco, CA, USA, 770 p.

Hastie, T. and Tibshirani, R. 2001. Classification by Pairwise Coupling. Annals of Statistics 26-2, p. 451-471.

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., Murthy, K. R. K. 2001. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13-3, p. 637-649.

Koliadin, V. 2012. Statistical Glossary. Available at: http://www2.statistics.com/resources/glossary/index.php3, 07.01.2013.

Lee, J. W., Lee, J. B., Park, M., Song S. H. 2005. An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics & Data Analysis,* 48-4, pp. 869-885.

Lu, Y.and Han, J. 2003. Cancer classification using gene expression data. Information Systems 28- 4, pp. 243-268.

Manning, C. D., Raghavan, P., Schütze, H. 2008. *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK, 544 p.

Platt, J. 1998. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In: B. Schoelkopf and C. Burges and A. Smola (eds), *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA, USA: The MIT Press, 386 p.

Polaka, I., Tom, I., Borisovs, A. 2010. Decision Tree Classifiers in Bioinformatics. *Scientific Journal of RTU*. 5th series, *Datorzinatne* 44, pp. 118-123.

Rokach, L. and Maimon, O. Z. 2008. Data mining with decision trees: theory and applications. Singapore: *World Scientific*, 244 p.

Comparison and evaluation of pattern matching algorithms for intrusion detection

Martins Sapats, Nauris Paulins

Latvia University of Agricultur, Liela Str. 2, Jelgava, LV-3001, Latvia
martins.sapats@llu.lv, nauris.paulins@llu.lv

Abstract: *In nowadays Network security is primary concerned in large organizations. Intrusion Detection Systems have become widely recognized as powerful tools for identifying, deterring and deflecting malicious attacks over the network. Essential to almost every intrusion detection system is engine ability to search through patterns database and identify content that matches known attacks, often it takes most of the processing time. One of innovations in this area is multi-core processors and multithreaded application design which is implemented in Next Generation Network Intrusion Detection System Suricata. For many years most popular Intrusion detection system was Snort, but it not using multithread design, however there are some research to implement this feature in Snort. However each system uses different pattern matching algorithms. In this paper we compare and evaluate the pattern of matching algorithms performance in these systems when booth Snort and Suricata has ability to use multithreaded design. Aim of this analysis is to test, is it Suricata algorithms more effective than Snort algorithms in multithreaded computing approach.*

Keywords: intrusion detection, network security, pattern matching.

Introduction

Intrusion detection system (IDS) is one of basic addition tools for network security infrastructure, which has ability to read network packets, captured from the network to identify attacks. There is two main detection techniques adopted in real production: misuse detection and anomaly detection. Misuse detection recognizes intrusion by matching collected data with pre-defined patterns of intrusive behavior. Anomaly detection hypothesizes that abnormal behavior is different from normal behavior, it learns normal behavior modes and detects anomaly by noticing deviations from normal behavior. Misuse detection systems are more popular in real enterprise networks, but anomaly detection is more popular in research area, because misuse detection can recognize just known attacks, but anomaly systems are able to recognize previously unknown attacks (Garciateodoro et al., 2009).

Commercial IDS often are very expensive so lot of companies searching for open source alternatives, which could be used for free. Snort has become most widely used network IDS worldwide, it is publicly available open-source IDS. It is PC-based IDS, but it has many third-party solutions, plugins and it is very popular in research area as well. Snort is currently written as a single-threaded user-level application. It utilizes the TCP/IP stack to capture and inspect packet payloads in order to identify signatures of malicious activities as shown in Fig. 1.

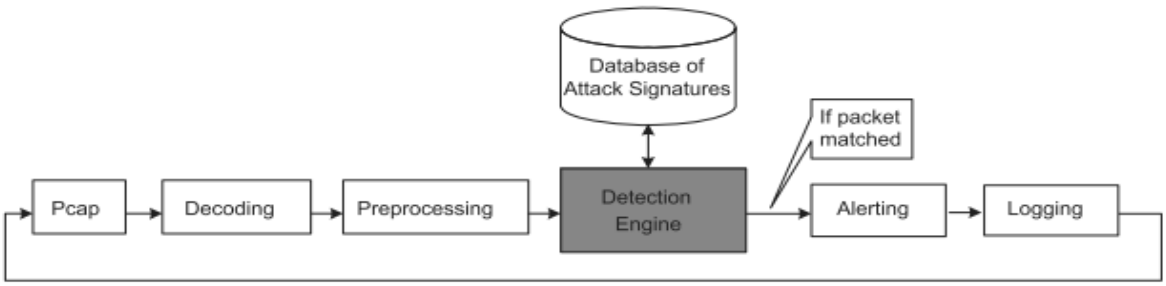


Fig. 1. **Basic components of Snort IDS.**

Typically, IDS resides on the edge of a network and performs deep packet inspection on every packet that enters the protected network against several thousands of attack signatures. The signatures are represented as a set of rules that are frequently updated by the security community (Salah and Kahtani, 2010). Detection process and signature checking process is very slow, it takes about an 80% CPU resources and as more rules system has to check as more time it takes. Systems are coming more powerful and CPU has multiple cores what makes them faster, but multi-core architecture is effective if software can support multithreading processes. There is some research done to increase Snot performance, like parallelization or pattern matching algorithm modifications, but originally it still working with old single thread design and old algorithms.

To solve shortcomings which has Snort, Open Information Security Foundation (OISF) has been released Suricata in 2009, which is declared as next generation intrusion detection system. Suricata offers multi-threading

and lot of other new features, like Automatic Protocol detection, Flow variables, better rule writing language features and Fast Internet Protocol (IP) matching (OISF, 2012). Suricata also is open-source misuse based system and because of new possibilities could change Snort which is released in 1998. Design of Suricata required development of original detection algorithms. The multi-pattern-matcher (MPM) is a part of the detection engine within Suricata that searches for multiple patterns at once. Generally, signatures have one or more patterns. Of each signature, one pattern is used by the multi-pattern-matcher. An example based on four cores is shown in Fig. 2. That way Suricata can exclude many signatures from being examined, because a signature can only match when all its patterns match.

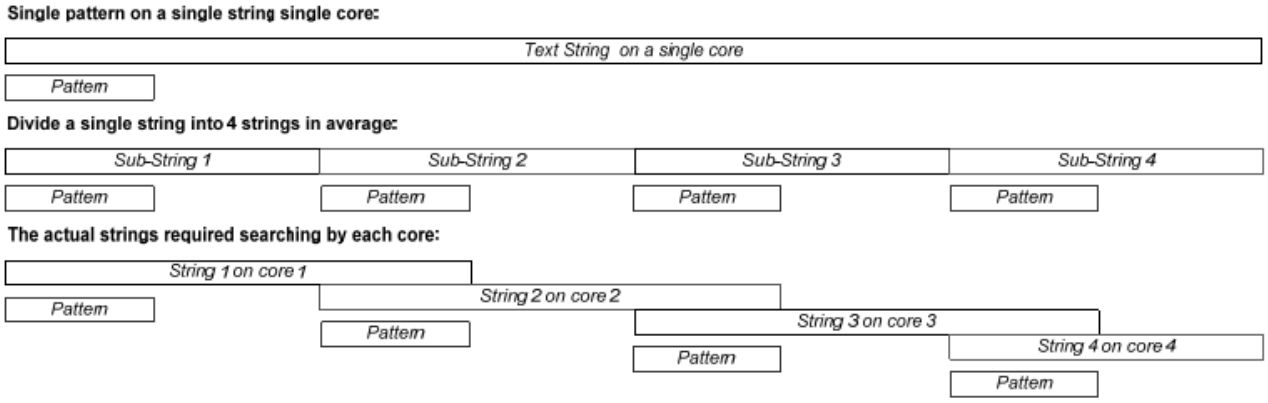


Fig. 2. **Multi-pattern-matcher on 4 processor cores.**

Aim of this paper is to compare Snort and Suricata performance in network security. This research will focus hot the performance and reliability of the IDS changes when configuration and the amount of network traffic is changed. Wouldn't be honest to compare Snort and Suricata performance on multiple cores. Snort has been modified to support process parallelization. Main interest in performance measurement will be in number of dropped packets and less on accuracy.

Materials and methods

The test-bed for IDS evaluation was constructed virtual platforms using Oracle VirtualBox. (Oracle, 2012). It is necessary for better initialization of experiment, because of several scenario re-configuration and repetition. As base system was chosen CentOS 6.3 64-bit Linux which is widely used in several servers. Hardware configuration was 3.3GHz Intel Core i3-3220 CPU with 8 GB RAM and 1 TB 7200RPM SATA HDD. Both systems use the same VRT rule sets and also the same logging methods – MySQL and Barnyard2.

For testing systems it was necessary to use datasets with collected traffic. It could be possible to use captured traffic in university, but it is not analyzed, so it is hard to evaluate results after experiments. It has been problem for last years to get datasets for traffic evaluation, because most popular datasets DARPA (MIT Licoln Labs, 1999) and KDD'99 (KDD-cup data set, 1999) was made more than ten years ago when computers was under systems like Windows 98. Finally last year Information Security Centre of Excellence has released new datasets, they have captured and carefully analyzed four weeks of network activity associated with the users and servers of our research centre for any abnormal or malicious activity. (Shiravi et al., 2012). Network traffic was replied on the network at different speeds, it was done using TCPReplay.

Before this research was evaluated several metrics (Paulins and Sapats 2012; Antonatos et al., 2004) for system testing, but this time we focus to compare multithreading capabilities and several algorithms used in systems, so just separate criteria was taken – number of dropped packets, CPU and memory workload, network bandwidth and total amount of alarms. System resources were monitored using Linux command line utility dstat.

Experiment was designed to provide data regarding how each systems basic algorithms working with increasing network speed and rule amount. Snort officially uses Aho-Corasick (AC) for multiple pattern matching. (Baker and Esler, 2007). However there are several implementations of these algorithms of snort:

- AC: Aho-Corasick Full (high memory, best performance);
- AC-BNFA: Aho-Corasick Binary NFA (low memory, high performance);
- AC-LOWMEM: Low Memory Keyword Trie (low memory, moderate performance);
- AC-SPLIT: Aho-Corasick Full with ANY-ANY port group evaluated separately (lowmemory, high performance).

Suricata also uses Aho-Corasick, but it uses some more Multi-Pattern-Matcher (MPM) algorithms - b2g, b3g, wumanber, ac. (Suricata.yaml, 2012) The falowing algorithms are:

- B2G – Implementation of the SBNDMq pattern matching algorithm that tries to be very limiting memory use and CPU cache efficient. This is revised version of SBNDM applying q-grams. (Durian et al., 2008).

- B3G – This is 3 gram implementation of the SBNDMq pattern matching algorithm.
- Wumanber – this is implementation of Wu-Manber pattern matching algorithm. (Zhang et al., 2009).

Each algorithm from both systems was tested with two different rule sets – 507, 3253 and also with two different network speeds – 250 Mbps and top speed (best speed what is possible to get from network interface).
As shown in Fig. 3 is placed offline of regular traffic and system is sniffing span port traffic in switch. Network card, which facing analyzed traffic, has no IP address. It means that IDS is more difficult to compromise from regular network. It has internal non-routable IP and access to open ports will be limited to administrative workstation. Such topology also provides more accurate measurement and has no impact from regular traffic.

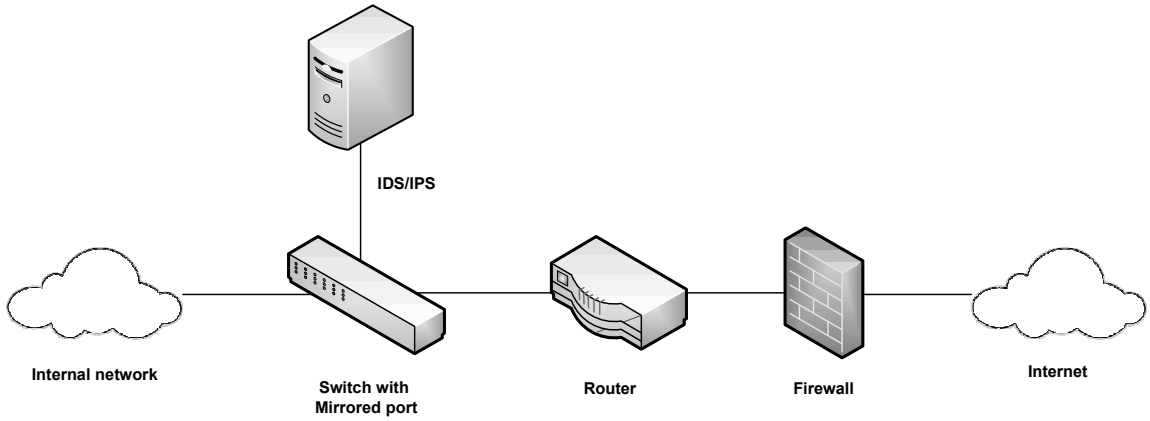


Fig. 3. **Experiment logical network diagram.**

The first we installed and run each detection engine individually in the virtual machine environment with the combined ET and VRT rule sets. Collect was used to record the CPU, RAM, network use and dropped packets of the server. Experiment was conducted separately on each system.

Results and discussion

In this section we detail overall results of measurements. Both Snort and Suricata uses Aho-Corasick as default multi pattern matching algorithm. So at first step Aho-Corasick was evaluated on both systems, to compare system performance with multithread design. This algorithm is implemented in both systems, so it is easy to compare system performance. Rest of the algorithms is shown on separate charts, so that it is possible to compare system CPU load and time for packet processing.
In first scenario both systems was loaded with 507 rules. Fig. 4. show that Snort using more CPU power, but as network traffic speed growth than Suricata is starting use more CPU than Snort as it is shown in Fig. 5. With greater network traffic speed both systems started to drop more packets. With 250 Mbps Snort already had about 23.8 % of dropped packets, but Suricata just 0.5 %, but with 700 Mbps Snort had a ready 33.9 % of dropped packets and Suricata had 15.9%. Also we noticed that with grater traffic Suricata used more RAM than Snort, where Suricata used about 3900 Mb RAM, but Snort about 3000 Mb. These results show that Suricata has bigger performance with default AC algorithm than Snort.

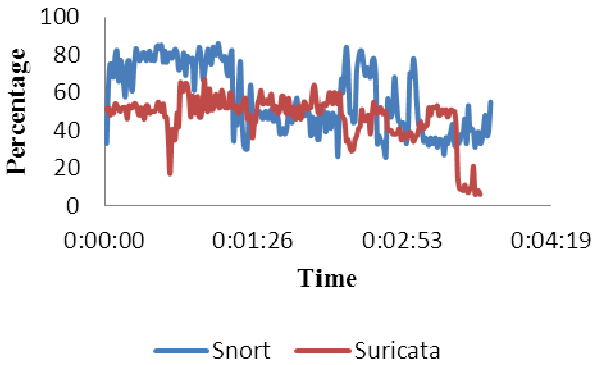


Fig. 4. **CPU load with AC, ~250 Mbps.**

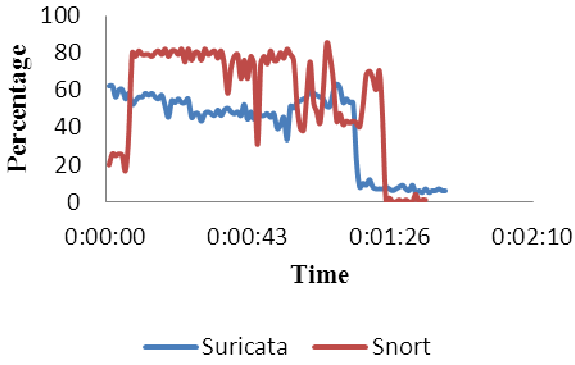


Fig. 5. **CPU load with AC, top speed ~700 Mbps.**

In case of Snort AC-BNFA and AC-LOWMEM algorithm performance was very similar as shown in Fig. 6. Also dropped packet percentage and memory usage was quite similar where AC-BNFA 29.59% and 2820 MB, but AC-LOWMEM 27.99% and 2970 MB. And even with greater network load AC-BNFA had smaller memory

usage than AC-LOWMEM. As you can see in Fig. 6. and Fig 7. AC-SPLIT requires lot less CPU usage and it shown also small percentage of dropped packets which was 9.73% and 15.73 on greater traffic, but biggest difference is that this algorithm working lot slower.

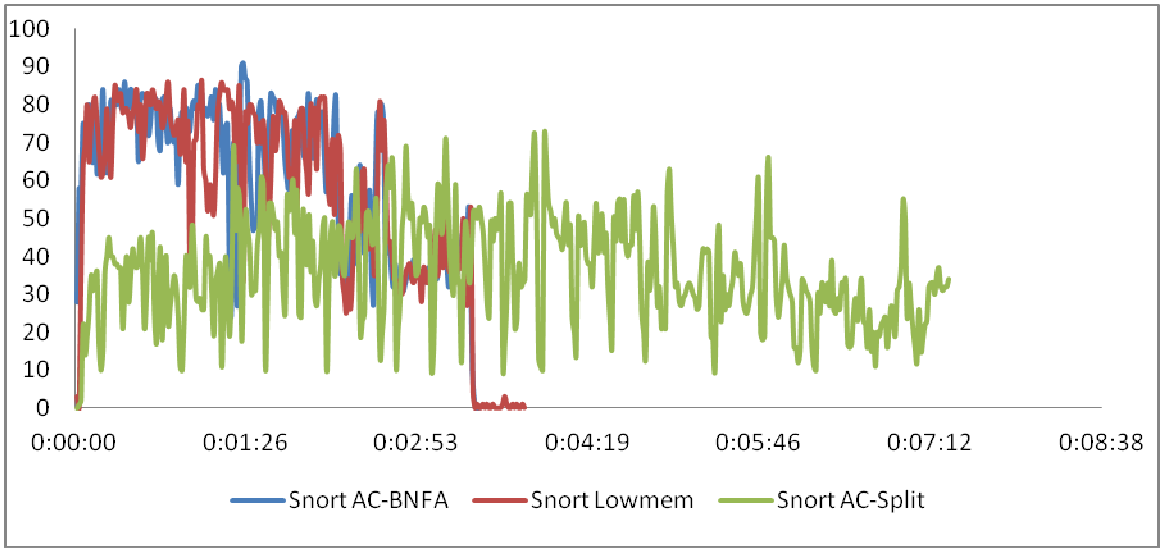


Fig.6. **Snort algoritm performance, ~250 Mbps.**

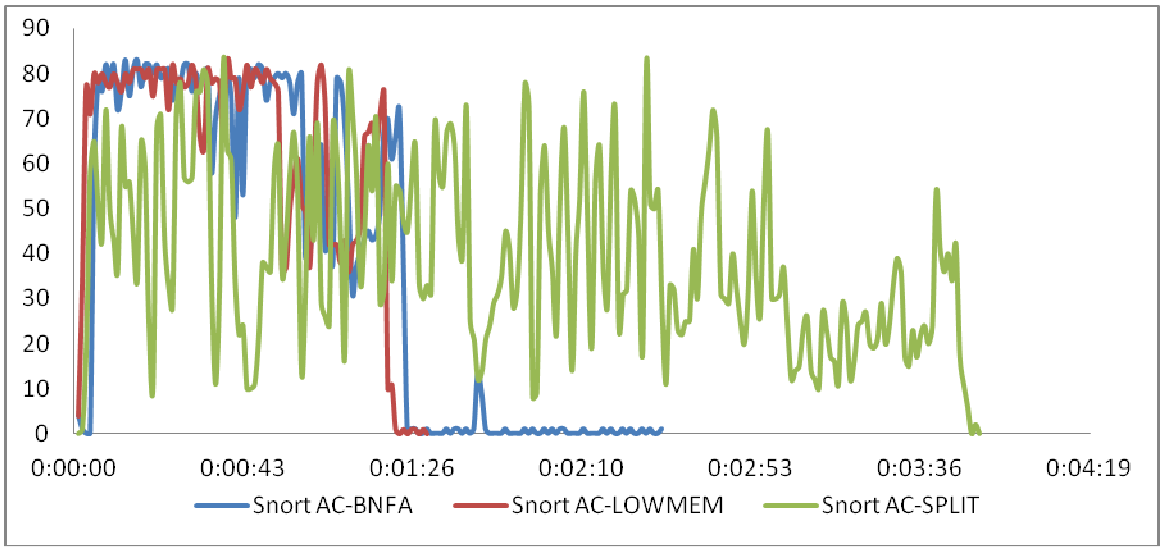


Fig.7. **Snort algoritm performance, ~700 Mbps.**

In case of Suricata performance is better, CPU load is not greater than 70% as you can see in Fig. 8 and Fig. 9. Suricata Wumanber taking less CPU power than others, however they are acting quite similar. Even B2G is slower it has better rate on dropped packets, on top speed it dropped just 1.5 % of packets where B3G had 28.3% and Wumanber 0.7 %. Suricata performance looks better than in case of Snort.

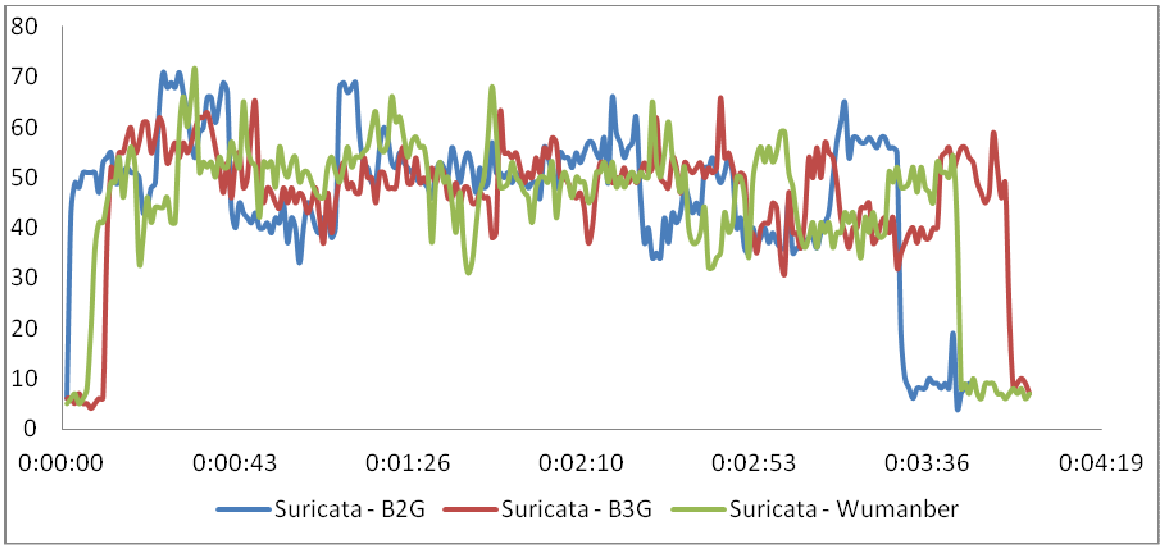


Fig.8. **Suricata algoritm performance, ~250 Mbps.**

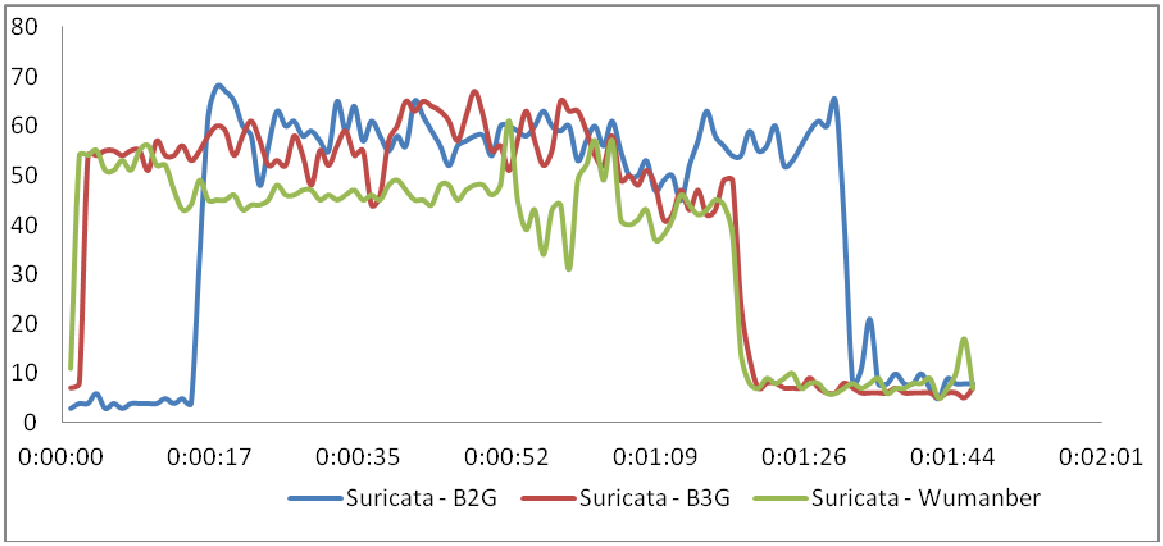


Fig.9. **Snort algoritm performance, ~700 Mbps.**

In second scenario both systems was loaded with 3523 rules, which should take more computational resources and more time to check each packet. As it is shown in Fig. 10 and Fig. 11 than CPU load did not increase very much and similarly like previous scenario Snort starting take more CPU when increase network traffic. Interesting that dropped packets in lower speed was not at all, but on top speed dropped packet percentage was just 8 % for Snort and 1% for Suricata. Also RAM usage did not increase more than previously.

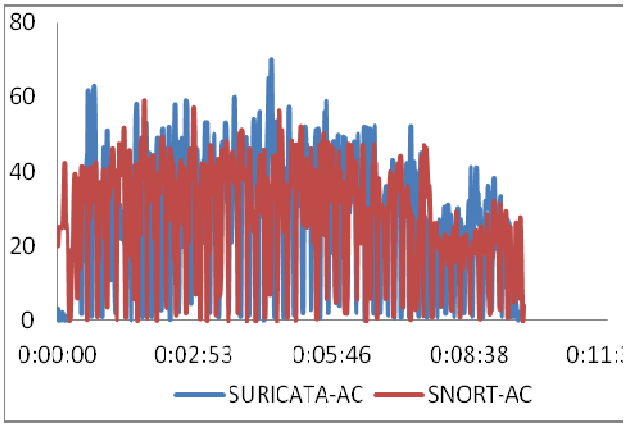


Fig.10. **Snort algoritm performance, ~250 Mbps.**

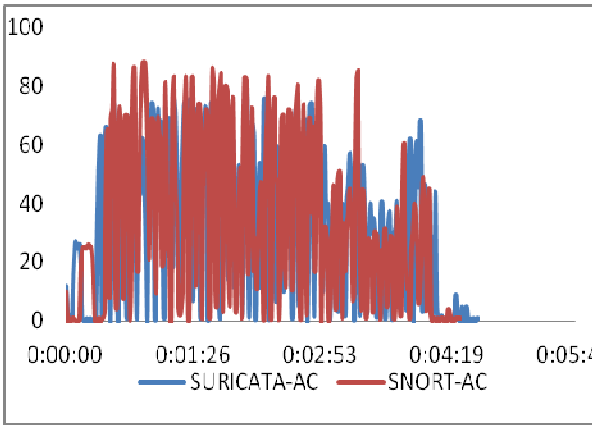


Fig.11. **Snort algoritm performance, ~700 Mbps.**

According to the previous scenario results, other algorithms were analyzed just on top speed, because there are no big changes in system load. Again Suricata is taking less CPU load as it is shown in Fig. 12 and Fig. 13. Suricata is not going up to 80%, but snort is taking about 90%. Suricata has better rate on dropped packets as well, it had 7.7% for B2G, 0.3 % for B3G and 0.7 % for Wumanber. Snort dropped packet rate was much bigger it had 16.23% for BNFA, 14.35 % for AC-SPLIT and 28.100 % for LOWMEM.

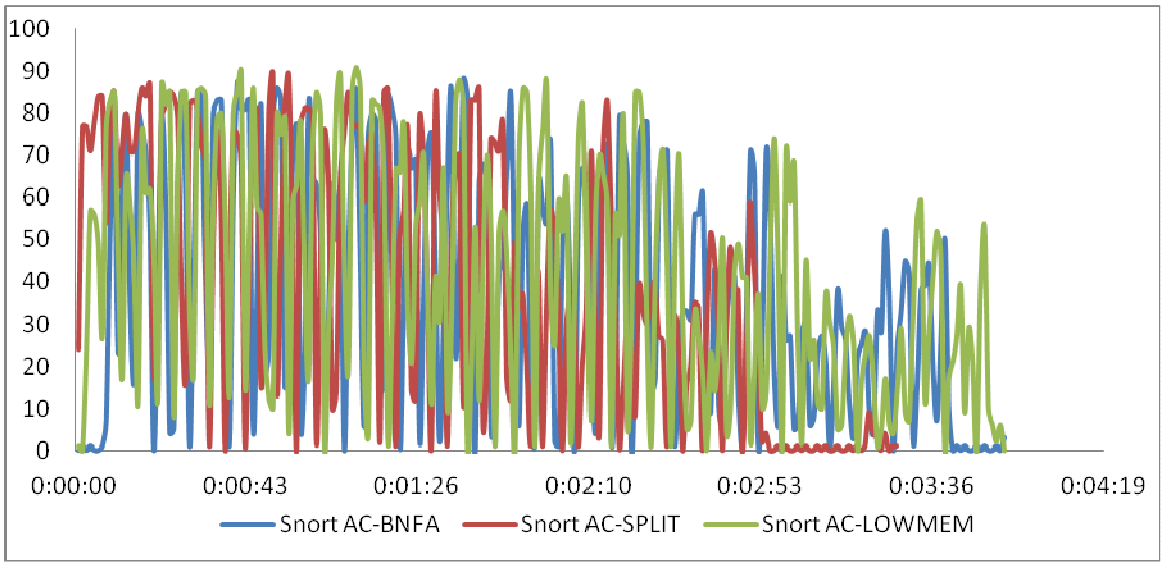


Fig.12. **Snort algoritm performance, ~700 Mbps.**

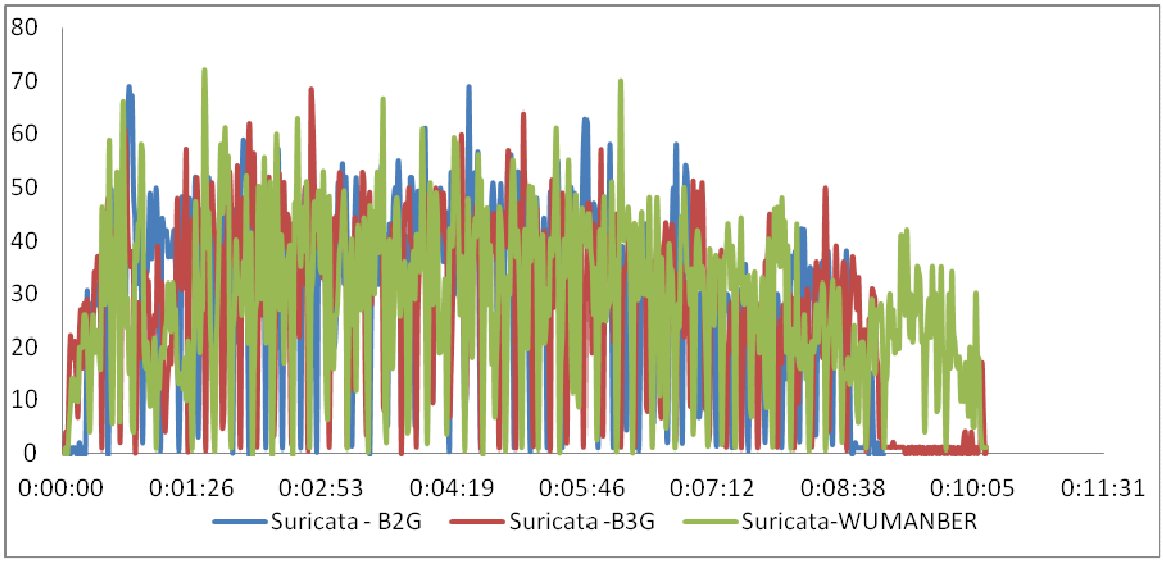


Fig.13. **Suricata algoritm performance, ~700 Mbps.**

Conclusion

This paper provides tow open-source network-based IDS system algorithm analysis. Snort is currently the de-facto standard for open-source network-based intrusion-detection systems around the world. Suricata is still in early stages of development but offers speed improvements and capabilities unavailable in Snort. The results of research have shown that Suricata has better performance in networks where traffic is more active and it has smaller percentage of dropped packets. Interesting option of results was that percentage of dropped packets was smaller in second scenario, we can't explain it yet, but this result we will evaluate in future experiments.

In both scenarios – with 507 rules and 3523 rules system load was very similar which means that systems are quite adaptive to such amount or rules? In further research we want to test system load and algorithm performance with bigger rule sets. As it is shown in charts, some algorithm effectiveness is possible to see just in bigger rule sets. AC algorithm is very effective, but on more active traffic and with more rules new generation hash algorithms are more effective and more accurate.

We tried to stay on default parameters on systems, and more concentrate on algorithm testing, but there are lot of parameters which can increase system performance, for example buffer size or packet size. Therefore it is

important to investigate other possibilities to improve IDS performance in traffic analysis which is or next investigation topic in future work.

Acknowledgements

This research has been supported by the European Social Fund within project "Support for the Implementation of Doctoral Studies at Latvia University of Agriculture", /2009/0180/1DP/1.1.2.1/09/IPIA/VIAA/017
Research has been developed with support of European Social Fund activity "Support to Master studies" project "Support to the implementation of LLU masters". Contract No. 2011/0020/1DP/1.1.2.1.1/11/IPIA/VIAA/011.

References

Antonatos, S. and Anagnostakis, K. G., 2004. Generating Realistic Workloads for Network Intrusion, Proceedings of the 4th ACM workshop on software and performance, ACM, pp. 207-215.

Baker, A. R. and Esler, J., 2007. Snort IDS and IPS Toolkit, Syngress Publishing, Inc, Burlington, 728 p.

Durian, B., Holub, J., Peltola, H., Tarhio J., 2008. Tuning BNDM with q-Grams. Proceedings of the Workshop on Algorithm Engineering and Experiments, pp. 29 – 37.

Garciateodoro, P., Diazverdejo, J., Maciafernandez, G., Vazquez, E., 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computer and Security, vol 28., pp. 18 – 28.

KDD-cup data set. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, 12.12.2012.

MIT Lincoln Labs, 1999. DARPA intrusion detection evaluation. Available: http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/, 12.12.2012].

OISF, 2012. Suricata. Opens Source IDS/IPS/NSM engine. Available: http://suricata-ids.org/features/, 11.12.2012.

Oracle, 2012. Virtualbox. Available: https://www.virtualbox.org/, 12.12.2012.

Paulins, N. and Sapats, M., 2012. Evaluation Methods of Network Intrusion Detection System , Proceedings of 5-th international scientific conference "Applied Information and Communication Technologies", pp. 27 – 33.

Salah, K. and Kahtani, A., 2010. Performance evaluation comparison of Snort NIDS under Linux and Windows Server. Journal of Network and Computer Applications, vol. 33, pp. 6 – 15.

Shiravi, A., Shiravi, H., Tavallaee, M, Ghorani, A., 2012. Toward developing a systematic approach to generate benchmark dataset for intrusion detection. Computer and Security, vol. 31, pp. 357 – 374.

Suricata.yaml, 2012. Available: https://redmine.openinfosecfoundation.org/projects/suricata/wiki/Suricatayaml, 12.11.2012.

Zang, B., Chen, X., Pan, X., Wu, Z., 2009. High Concurrence Wu-Manber Multiple Patterns Matching Algorithm. Proceedings of the 2009 International Symposium onf Information Processing (ISIP'09), pp. 404 – 409.

Improvement of neural networks learning by feature extraction methods

Sergejs Kodors

Rezekne Higher Education Institution, Brivibas street 39-9, Rezekne, LV-4601, Latvia
sk_7@inbox.lv

Abstract: *This paper discusses a comparison of the feature extraction methods, if a classifier of recognition problem is the artificial neural network with the back-propagation algorithm. The feature extraction methods can improve the classification accuracy and minimize a size of an education dataset and a signal processing time. All these improvements are satisfied by the transformation of the recognizable signal and by the minimization of the signal size. There are two problems to measure these improvements: the improvement of the recognition can be only determined in the experiment, the second problem is that a measured structure of the artificial neural network can contain the unlimited number of the layers and the unlimited number of the perceptrons in every layer. Therefore there is need to argument the chosen parameters of the experiment. The goal of this work is to organize the experiment plan to compare the feature extraction methods. The paper contains the description of the structure of the artificial neural network, the dataset and the elements which are influenced by the feature extraction methods.*

Keywords: artificial neural network, feature extraction, improvement.

Introduction

The artificial neural networks and the neurocomputers are a branch of the computer science, which is based on an imitation of the human brain behaviour. The first artificial neural networks and neurocomputers were described and proposed by the authors McCulloch, Pitts and Rosenblatt in the 1950s (Jasnickiy, 2005).
The artificial neural network is the distributed system of the computing elements – perceptrons. The perceptron imitates a structure and work of the neuron.
The neurocomputer is a type of the processor, which works using the principles of the artificial neural network.
The artificial neural networks are used in the different fields:

- The medical diagnoses;
- The lie detectors;
- The estimators of the exchange rate;
- Other fields.

The artificial neural networks have the following positive features (Haykin, 2006):

- Nonlinearity;
- Input-output mapping;
- Adaptation;
- Fault tolerance;
- Very-large-scale-integrated (VLSI) implementability;
- Uniformity of analysis and design.

The application of the artificial neural networks has some expenses to get the classification accuracy: a complexity and a size of dataset; where the complexity is the signal processing time and the complexity of the artificial neural network structure. These relations are illustrated in Fig. 1.

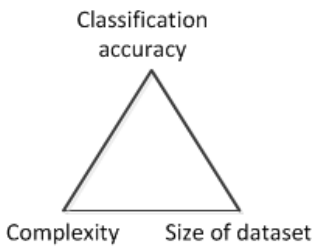


Fig. 1. **Relations of artificial neural network parameters**.

All these relations are explained by the VC-dimension.
The VC-dimension is the maximal number h of the vectors, which can be separated into two classes by 2^h possible ways. According to (Koiran and Sontag, 1996), the artificial neural networks, which use the sigmoidal activation function (1), have the VC-dimension at least as large as the square of the number of the weights.

$$\varphi(v) = \frac{1}{1 + \exp(-v)} \tag{1}$$

where φ – an activation function;
v – an induced local field.

This means that the huger dataset needs the more complicated artificial neural network to get the necessary accuracy. According to the curse of dimension, the function of the higher-dimensional space is rather the more complicating than the function of the lower-dimensional space. In other words, if a signal is converted from the higher-dimensional space to the lower-dimensional space, the classification accuracy is improved, but the complexity and the size of the dataset are reduced.

The feature extraction methods improve the artificial neural network classification accuracy and provide the faster and the more cost-effective classifier reducing the number of the signal features.

Feature extraction

The supervised learning is using a training set $S^m = \{x^i, y^i\}_{i=1}^m$ to train the artificial neural network, where y^i - an expected output, x^i - an input, where $x \in \Re^N$, so N coordinates are called features (Krupka et al., 2008).

The feature extraction is a technique to select the features from a transformed space (Li et al., 2009).

The feature extraction is used:

- To improve the classification;
- To reduce the number of the dimensions;
- To get the scale, rotation and translation invariant features.

The scheme of the feature extraction influence can be depicted as the following figure (Fig. 2):

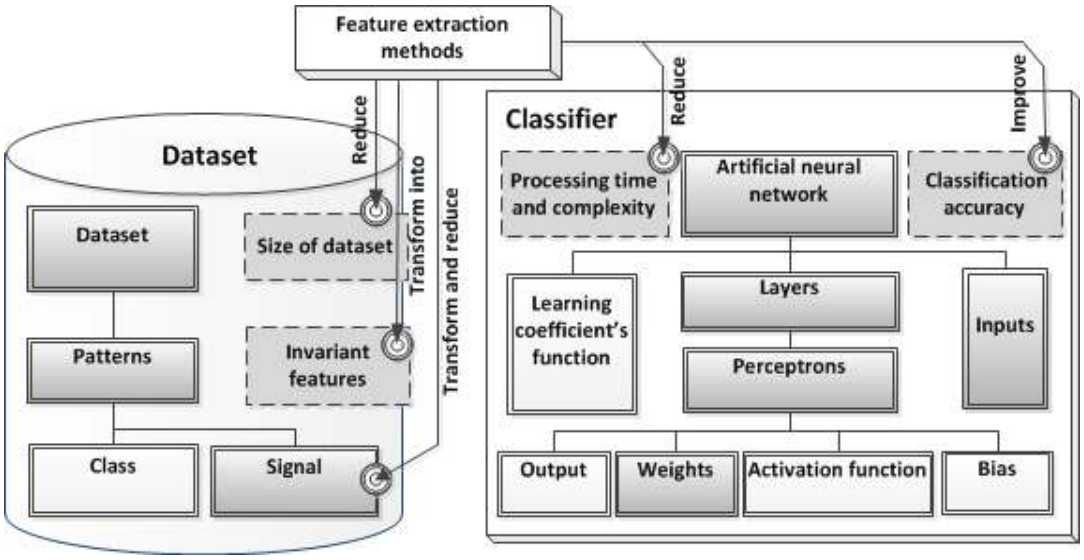


Fig. 2. **Influence of feature extraction methods**.

There are several examples of the feature extraction methods:

- Discrete Fourier transform (DFT);
- Wavelets;
- Principal component analysis (PCA);
- Independent component analysis (ICA);
- Fisher linear discriminant analysis (FLDA).

Back-propagation algorithm and feature extraction methods

The back-propagation algorithm was developed in 1986 (Haykin, 2006). It is the supervised learning algorithm to train the multilayer artificial neural network.

The back-propagation algorithm consists of two stages (Haykin, 2006; Jasnickiy, 2005; Nikolenko and Tulupjev, 2009):

- The forward pass – firstly the signal, which must be classified, is input into the artificial neural network, then it is being processed from a layer to next layer, until the output signal is generated. There are static weights and biases in this process;

- The backward pass - the output, which was generated by the forward pass, is compared with the expected output, the difference is calculated and sent back as an error signal to correct the weights.

In the training stage the feature extraction methods are used as the transformation algorithm of the dataset (Fig. 3) and as the preprocessing in the exploitation stage (Fig. 4).

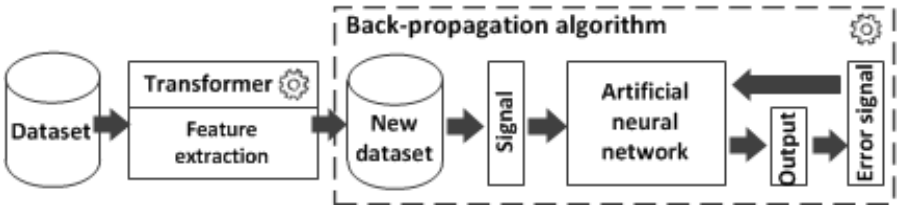


Fig. 3. **Training stage**.

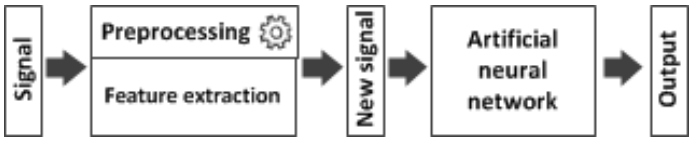


Fig. 4. **Exploitation stage**.

There is possible to see in Fig. 3 and Fig. 4, that every expense has its own field of the significance:

- The dataset is only used in the training stage, therefore the size of the dataset is not important in the exploitation stage;
- The complexity may be not so important in the training stage, but it is important in the exploitation stage.

Argumentation of experiment parameters

To design the experiment to measure the improvement by the feature extraction method, there is need to define the range of the experiment.

What is the minimal classification accuracy?

The minimal classification accuracy is the case, when there is possible to argue, that the correctly classified signals have not been guessed. Therefore there is need of the calculating the probability to guess the results and the method to verify, that the results are different enough.

So there are 2^n possible outputs, where n is a number of the output perceptrons, then there is the probability $1/2^n$ to guess the expected output.

The Chi-square test is a statistical measure, which is used to make comparison between the theoretical populations and actual data to test the goodness of fit. It can be applied to verify the difference for the artificial neural networks (Table 1).

Table 1

Chi-square test for artificial neural network

degree of freedom = 1	Correct answers	Wrong answers
Expected frequency	$p(A)\cdot N = \frac{1}{2^n}N$	$p(\overline{A})\cdot N = \left(1-\frac{1}{2^n}\right)N$
Observed frequency	m	$N-m$
Chi-square	$\chi^2 = \frac{(m-p(A)\cdot N)^2}{m} + \frac{((N-m)-p(\overline{A})\cdot N)^2}{N-m}$	
Minimal classification accuracy is satisfied	If $m > \frac{1}{2^n}N$ and $\chi^2 \geq \chi^2_\alpha$	
α - a level of significance, N – a size of the test set, n - a number of the output perceptrons, χ^2_α - the critical Chi-square, m – a number of the correct answers		

What is the minimal size of the test set?

The cross-validation method is used to measure the possibility of the artificial neural network to classify the signals, which were not used in the training. The cross-validation method divides the dataset on the training set and the test set:

- The training set is used in the learning process;
- The test set is used to measure the classification accuracy.

According to (Kothari, 2004), there are the following conditions for the application of the Chi-square test:

- The overall number of the items are at least 50;

- Every group contains at least 10 items.

These conditions can be applied to the test set. If the observed frequency is less than 10, there can be applied the following rules:

- If the number of the wrong answers is less than 10, the minimal classification accuracy is satisfied, because 40 from 50 is 80% , but the maximal expected number of the correct answers are 50% for the guess, when there are only one output perceptron;
- If the number of the correct answers is less than 10 and $N/2^n \geq 10$, then the minimal accuracy is not satisfied.

What is the minimal size of the dataset?
According to (Haykin, 2006), the ratio of the training set to the test set is 4:1. So, if the minimal size of the test set is at least 50 items, then the training set must be at least 200 items. It means the minimal size of the dataset is 250 items.
How long must be the artificial neural network trained?
The early stopping method of training uses the cross-validation method to determine the end of the training process. The artificial neural network has been trained for *Z* iterations and then it is tested by the test set. While the classification accuracy is rising, the training process is repeated.
What is the minimal number of the hidden layers?
The universal approximation theorem states that two-layer (with one hidden layer) artificial neural network with the sigmoidal activation function can approximate the classification function of the input-output set with the unspecified size (Haykin, 2006; Jasnickiy, 2005).
What is the maximal number of the perceptrons?
For an evaluation of the number of the perceptrons in the hidden layer, it is possible to use the formula (2), which is the result of the Kolmogorov-Arnold-Hecht-Nilsen theorem (Jasnickiy, 2005).

$$\frac{N_y Q}{1+\log_2 Q} \leq N_w \leq N_y\left(\frac{Q}{N_x}+1\right)\cdot\left(N_x+N_y+1\right)+N_y \tag{2}$$

where N_y – a number of the output perceptrons;
Q – a size of a training set;
N_w – a number of the weights;
N_x – a number of the inputs.

So there is possible to calculate the number of the perceptrons in the hidden layer of two-layer artificial neural network by the formula (3).

$$N = \frac{N_w}{N_x + N_y} \tag{3}$$

where N – a number of the perceptrons in the hidden layer.
What is the minimal number of the measures?
The central limit theorem states if a sample is from the normal population, the mean of this sample is itself normally distributed; if the population is not normally distributed, the shape of the distribution depends largely on the shape of the parent population, when a size of the sample is small; but as the size is getting larger ($s > 30$), the shape is becoming more and more like a normal distribution.
What is the value of the learning coefficient and the iterations?
The training process of the artificial neural network and the adaptation process of the self-organizing map have some similarities, therefore there is possible to use the recommendations for the self-organizing maps.
The source (Haykin, 2006) advices the following parameters:

- The learning coefficient $\eta \in [0.01;0.1]$;
- The number of the iterations to organize the map is equal to 1000;
- The formula to calculate the learning coefficient:

$$\eta = \eta_0 \cdot \exp(-\frac{n}{\tau}) \tag{4}$$

where n – an iteration;
η_0 – an initial learning coefficient;
τ – some coefficient.

- The recommended parameters for the formula (4): $\eta_0 = 0.1$ and $\tau = 1000$.

Results and discussion

All the argumentations and the ranges of the experiment, which were discussed in the previous section, can be formed as the following experiment design:

1. Prepare at least 250 items for the dataset;
2. Convert the dataset by the feature extraction method;
3. Randomly take at least 50 items for the test set from the dataset;
4. Calculate the maximal number of the weights using the formula (2);
5. Calculate the maximal number of the hidden perceptrons N_h using the formula (3);
6. Prepare the artificial neural network with one hidden layer and one hidden perceptron, the activation function is the sigmoidal function (1);
7. Train the artificial neural network using the early stopping method of training. The training time is 1000 iterations before the testing, the formula of the learning coefficient is (4), the initial learning coefficient is 0.1 and the coefficient $\tau = 1000$;
8. Repeat the 7th step 99 times and calculate the mean observed frequency of the correct answers;
9. If the mean observed frequency is greater than $1/2^n$, the difference is verified by the Chi-square test, otherwise go to the 11th step;
10. If the Chi-square test has showed that the observed frequency deviates from the expected frequency, output the classification accuracy and the related number of the hidden perceptrons, otherwise go to the 11th step;
11. If the number of the hidden perceptrons is less than N_h, increase the number of the hidden perceptrons by 1 and back to the 7th step, otherwise output the best result of the classification accuracy and the related number of the hidden perceptrons.

Conclusion

This work was prepared with a goal to compare different feature extraction methods for RGB image recognition. Using the described experiment design the scientist can define the minimal number of the hidden perceptrons to classify the dataset. The group of the parameters: the minimal number of the hidden perceptrons, the number of the features, the number of the output classes and the dataset; can form the baseline to compare the feature extraction methods. If two experiments have the same datasets, comparing the minimal number of weights to classify the dataset, one can compare the improvements of two feature extraction methods. If the number of the features are the same, one can compare the number of the perceptrons, otherwise there is need to calculate the weights by the formula (5).

$$W = (i+1) \cdot p_h + (p_h + 1) \cdot p_{out} \tag{5}$$

where W – a number of the weights;
i – a number of the features;
p_h – a number of the hidden perceptrons;
p_{out} – a number of the classes (output perceptrons).

This paper has not description of the complication of the dataset, but this parameter is important to compare the results in the case, if the different datasets were used in the experiments. There is need to add, that the described experiment is only useful for the huge populations, for example, the characters recognition with the image size equal to 3x5, then the population will be 2^{15} items, if there are only the black and white pixels.

References

Haykin, S., 2006. Neironye seti: polnii kurs, izdanie vtoroe (Neural networks: a comprehensive foundation, second edition), Viljams, Moskva, 1104 p. (In Russian)

Jasnickiy, L., 2005. Vvedenie v iskusstvennyi intellekt (Introduction to artificial intelligence), Akademija, Moskva, 176 p. (In Russian)

Koiran, P., and Sontag, E.D., 1996. Neural networks with quadratic VC dimension. Advances in Neural Information Processing Systems, 8, pp.197-203.

Kothari, C.R., 2004. Research methodology: methods and techniques, 2nd ed., New Age International Publishers, New Delhi, 418 p.

Krupka, E., Navot, A., Tishby, N., 2008. Learning to select features using their properties. Journal of Machine Learning Research, 9, pp.2349-2376.

Li, Y., Lu, B.-L., Zhang, T.-F., 2009. Combining feature selection with extraction: unsupervised feature selection based on principal component analysis. International Journal on Artificial Intelligence Tools, 18, pp.883-904.

Nikolenko, S.I. and Tulupjev, A.L., 2009. Samoobuchajushiesja sistemy (Self-instruction systems). Moskovskii center nepreryvnogo matematicheskogo obrazovanija, Moskva, 288 p. (In Russian)

Ontology method construction for intelligent decision support systems

Olga Burceva, Ludmila Aleksejeva

Department of Modelling and Simulation, Riga Technical University, 1 Kalku Street, Riga, LV-1658, Latvia
olgaa.burceva@gmail.com, l.aleksejeva@rtu.lv

Abstract: *The lessons learned from experiments with early rule-based systems were not only that domain knowledge should be represented in an explicit way (such as the one supported by an ontology), but also that the problem-solving behaviour of a system should be carved out in a separate component of the system. The difficulty of finding the best alternative is often complex to resolve, especially it increases when it is necessary to consider alternative by several qualitatively different criteria. Available at the enterprises many information systems and decision support systems (DSS) are not able to fully meet the needs of managers because usually they represent a set of disparate databases, and the information in such systems is represented by text, or in the form of various directories and is characterized by a set of independent, uncoordinated and an implicitly expressed conceptual description of the system. Due to this it is necessary to have DSS that would provide decision makers information about the functions and mutual relations of the structural components of the company for their effective management. There are different approaches, models, and knowledge definition language for this but more and more popular recently become ontology engineering. The main idea of our approach is to separate the decision making method (method ontology) from the database (domain ontology) in the construction of ontology-based decision support system. Here, the problem of constructing domain ontology and method ontology appears. This paper focuses on constructing only method ontology.*

Keywords: method ontology, ontology construction, multicriteria decision support, AHP

Introduction

The difficulty of finding the best alternative is often complex to resolve; especially it increases when it is necessary to consider alternatives by several qualitatively different criteria.
Many information systems and decision support systems (DSS) available at the enterprises are not able to fully meet the needs of managers because they usually represent a set of disparate databases, and the information in such systems is stored as text or in the form of various directories and is characterized by a set of independent, uncoordinated and implicitly expressed conceptual descriptions of the system (Bolotova, 2012). Due to this, it is necessary to have DSS that would provide decision makers information about the functions and mutual relations of the structural components of the company for their effective management.
There are different approaches, models, and knowledge definition languages for this but ontology engineering is recently becoming more and more popular. At the formal level ontology is a system containing certain concepts, properties of concepts, relationships between concepts, and additional limitations as are determined by axioms (Khakhalin, 2005). Ontological engineering includes: definition of the concepts in the ontology, guidance taxonomy, development of concepts and situations structures, determination of properties and values of these properties, procedures for output and transformations of situations.
Ontologies provide a number of useful features for intelligent systems, as well as for knowledge representation in general for the knowledge engineering process. Ontologies perform an integrating function, providing a common conceptual basis in the decision-making processes and a common platform to bring together a variety of information systems.

Decision support systems

The main problem of the decision-making theory is the selection of one or more best objects (options, alternatives, etc.), ordering or ranking objects based on their properties, classification or sorting objects by the specified categories. The properties of these objects are characterized by many attributes or estimates on many criteria, the available quantitative and / or qualitative scale (Aleksejeva, 1998).
Preferences of the decision-maker (DM) are a key factor of rational choice. DM formalizes preferences, setting characteristics of the researched problems and properties of the objects, comparing solutions, evaluating the quality of the selection. Preferences can be defined by binary relations, functions, decision rules that have logical, mathematical and verbal form. At the same time, solving the problem, a person can express his preferences consistently. There are a lot of methods for selecting the objects described by many quantitative and/or qualitative attributes (Aleksejeva, 1998, Saaty, 2011). The best options selection is carried out using a variety of methods of optimal choice based on the search of one extreme or the many features that characterize the performance or quality of the solution (Petrovsky, 2009[a]). In the methods of multi-criteria optimization, generalized criterion is usually given by the convolution of many private numeric criteria in the form of a "weighted sum". However, determination of baseline weight is a serious problem.

To order objects in general or by multiple criteria, there are commonly used methods based on pairwise comparisons of objects. If there are many criteria and/or multiple decision makers, the resulting ordering of objects is constructed based on the pairwise comparisons of estimates of vectors representing the objects. In the methods of analytical hierarchy (Saaty, 2011; Petrovsky, 2009[b]) the options are ranked according to their priority, which is consistently evaluated by pairwise comparison of options, assessment criteria and participants to the global goal of the problem being solved. This method will be used in this work to show, how the method ontology can be constructed.

The Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process was devised by Thomas L. Saaty (1977) in the early seventies. It is a powerful and flexible tool for decision-making in complex multi-criteria problems. This method allows one to gather knowledge about a particular problem, to quantify subjective opinions and to force the comparison of alternatives in relation to established criteria (Saaty, 2011; Loranzo-Tello et al, 2008). The AHP method includes the following steps:
Step 1: making the hierarchy. The problem should be defined in a hierarchical structure. The hierarchical structure is like an inverted tree. At the top should be the goal to be achieved, or the problem to be solved. The following are the parameters which affect the value of the final decision – criteria. It should be noted that the criteria can be split into subcriteria. The next are alternatives to achieve the goal. For each of these alternatives it must be possible to determine the absolute and relative importance of each criterion. Thus, the hierarchy allows us to decompose a complex problem into parts, which allows us to understand the complexity and diversity of the upcoming elections (Saaty, 2011).
Step 2: setting priorities. All the criteria by which we are going to compare the alternatives must be mutually compared.
Step 3: comparison of the alternatives. With knowledge of the relative importance of each criterion, we can go to the comparison of alternatives for each criterion.
Step 4: check for consistency. If the procedures described above are performed by a group of DM, it is important to use the average of the personal ratings.
Step 5: making the final decision. With the results for the pairwise comparison of alternatives and the relative importance of the criteria, we can calculate the evaluation of each of the alternatives, which will give us a basis for making the final decision.

Ontology engineering

Ontology defines the common words and concepts used to describe and represent an area of knowledge, and so standardize the meaning. Ontologies are used by people, databases, applications that need to share domain information. Ontologies include computer usable definitions of basic concepts in the domain and the relationships among them. They encode knowledge in a domain and also knowledge that spans domains. So, they make that knowledge reusable (Rothenfluh et al, 1996). At the formal level ontology is a system containing certain concepts, properties of concepts, relationships between concepts, and additional limitations as are determined by axioms (Khakhalin, 2005). Ontological engineering includes: definition of the concepts in the ontology, guidance taxonomy, development of concepts and situations structures, determination of properties and values of these properties, procedures for output and transformations of situations.
There are many reasons why the need of ontology development appears (Grechko, 2005):

- for knowledge sharing among people or software agents total understanding the structure of the data;
- for re-use of knowledge of the subject area;
- to turn assumptions in explicit connection or dependence;
- to separate domain knowledge from the operational knowledge;
- to analyze the domain knowledge.

Ontological engineering denotes a set of design principles, development process and activities, supporting technologies, and systematic methodologies that facilitate ontology development and use its life cycle-design, implementation, evaluation, validation, maintenance, deployment, mapping, integrations, sharing, and reuse.

Related works

There are some definitions of methodology for building ontologies, again assuming manual approach. For instance, the methodology proposed in (Uschold et al, 1995) involves the following stages: identifying the purpose of the ontology (why to build it, how will it be used, range of the users), building the ontology, evaluation and documentation. The building of the ontology is further divided into three steps. The first is ontology capture, where key concepts and relationships are identified, a precise textual definition of them is written, terms to be used to refer to the concepts and relations are identified, the involved actors agree on the definitions and terms. The second step involves coding of the ontology to represent the defined conceptualization

in some formal languages (committing to some meta-ontology, choosing a representation language and coding). The third step involves possible integration with existing ontologies.
Traditionally, ontologies for a given domain are constructed manually using some sort of languages or representation and rely on manual extraction of common sense knowledge from various sources. Recently, several programs that support manual ontology construction have been developed, for example, METHONTOLOGY (Ferndndez et al, 1997) or Protégé (Youn et al, 2006).
Protégé 2000 assumes that knowledge-based systems are usually very expensive to build and maintain because knowledge-based system development is done by a team including both developers and domain experts who may be less familiar with computer software. Protégé 2000 guides developers and domain experts through the process of system development. Developers can reuse domain ontologies and problem-solving methods with Protégé 2000, shortening the time for development and program maintenance. One domain ontology that solves different problems can be used in several applications, and different ontologies can use the same problem-solving methods (Youn et al, 2006).
The PROTÉGÉ-II is used in (Rothenfluh et al, 1996). There is shown also how reusable domain and method ontologies are combined into the task-dependent application ontology. The construction of a knowledge-based system starts from a declarative description of the domain and of the problem-solving method: the domain and method ontologies. The developer merges these ontologies to produce an application ontology that is both domain- and method-specific. To generate a run-time system, PROTÉGÉ-II interprets the knowledge base created by the expert as input to the problem-solving method.
According to (Liu et al, 2005), decision ontology can be designed and used to conceptualize the knowledge for decision making process. Ontology can be used for the model base design and model management. Ontology is divided into two parts, i.e. domain ontology and modelling ontology, where domain ontology shows the terminology for decision making, concepts and terms. Decision making ontology which has been built in (Kornyshova et al, 2005) also is representing decision making knowledge; it includes concepts, their properties and relationships. Therefore, using the decision ontology the domain knowledge and decision models become more sharable and reusable to users from different background and interoperable to other software agents (Liu et al, 2005).

Suggested approach

The main idea of our approach is to separate the decision making method (method ontology) from the database (domain ontology) in the construction of ontology-based decision support system. The method ontology describes domain-independent method concepts, in contrast to the domain ontology, which describes method-independent domain concepts. Method ontologies are abstract descriptions of the inputs and outputs of the problem-solving method. The method ontology describes the knowledge requirements and the knowledge roles of a given problem-solving method (Rothenfluh et al, 1996).
Knowledge about the given domain may lead the developer to make changes in a generic method through method configuration. A generic problem-solving method should be decomposable: it should be divisible into some sequence of subtasks, which in turn are solved by other methods (Rothenfluh et al, 1996).

Tasks to be solved

In this work we are building the ontology method for decision support algorithm – AHP. Fig. 1 shows a flow-chart of the decision making process construction as logical basis for ontology construction.
The flow-chart represents all general steps of AHP method. Using the given flowchart with it steps, the method ontology will be constructed in the next chapter.

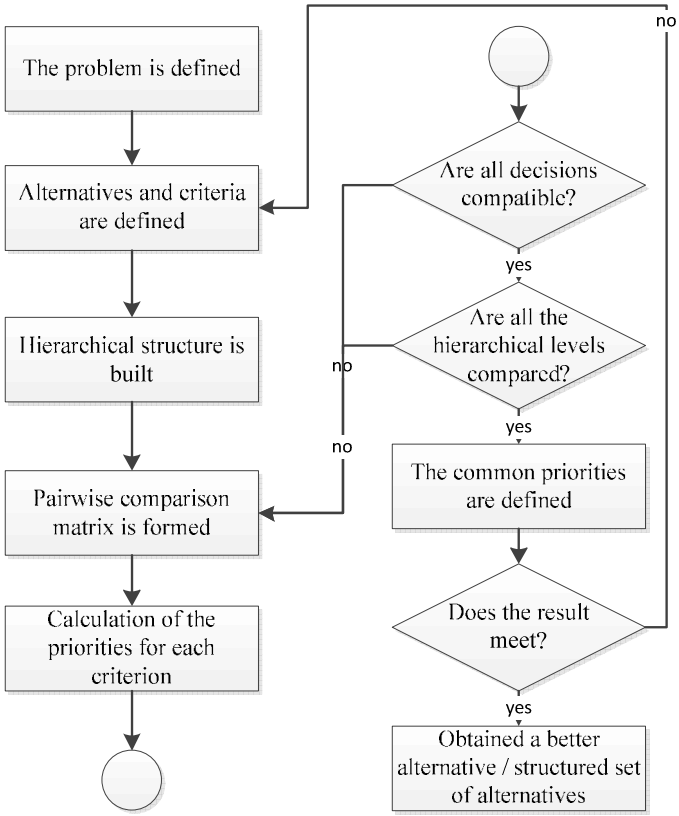


Fig. 1. **Flow-chart of the decision making process in AHP.**

Implementation of the suggested idea

In this chapter a case study of the AHP method ontology construction is described. The first step of the AHP method is to define the problem and the main objective is to make the decision. A first concept of ontology hierarchy can be a problem class.

The next step of the method is a hierarchy tree building where the root node is a problem, the intermediate levels are the criteria, and the lowest level contains the alternatives. In the hierarchy there are also clusters. Cluster is a group of nodes at the same levels which are subordinated to some other level – the top of the cluster. The clusters are formed by placement of links between nodes. So, the hierarchy has clusters and levels, in which one there are nodes of hierarchy – goal, criterion and alternative (Fig. 2).

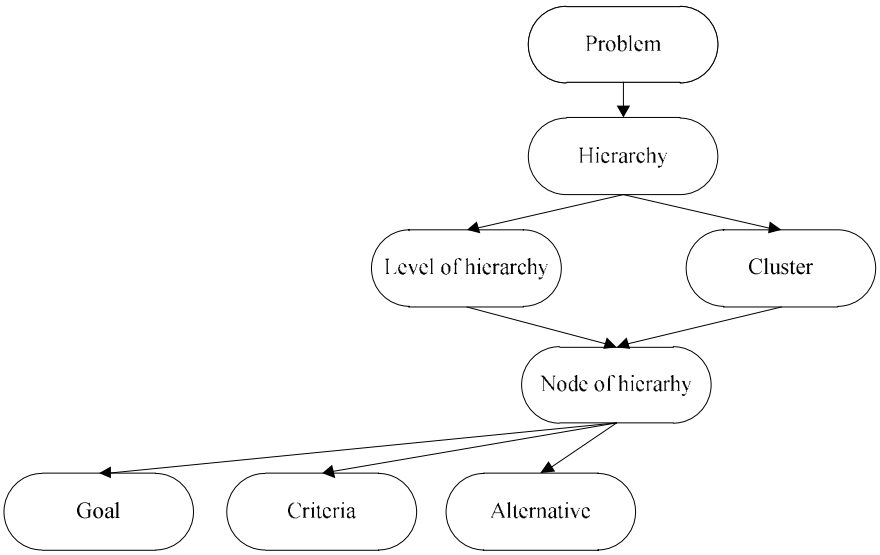


Fig. 2. **Fragment of the method ontology.**

There can be many final results of the problem– a set of alternatives, arranged alternatives, best alternative. Those all will be defined as the next nodes of the goal. To construct our method ontology, we will further take a

simple task which we will attempt to resolve using AHP method and at the same time constructing the next concepts of ontology.

We will take decision matrix with four alternatives and six attributes with numerical values (Table 1). The next step of the AHP method is evaluating the hierarchy. Once the hierarchy has been constructed, the building of pairwise comparison matrix for each level must be done. It is necessary pairwise compare all the criteria by which we are going to compare the alternatives.

Table 1

Decision matrix

	X1	X2	X3	X4	X5	X6
A1	2.0	1.5	2.0	5.5	5	9
A2	2.5	2.7	1.8	6.5	3	5
A3	1.8	2.0	2.1	4.5	7	7
A4	2.2	1.8	2.0	5.0	5	5

In the beginning, it is necessary to get the criteria evaluation. We evaluate the criterion mutual influence, using the ratio of the relative importance of the nine-point scale where 1 – equal importance, 2 – very slight superiority, lightweight superiority. Suppose that X1 criterion is more important than the X2 criterion with very slight superiority. As the matrix is symmetric, then X2 is better than X2 in ½ times. In this way, all the criteria are compared with each other and the pairwise comparison matrix is built; it can be seen from Table 2. Taking into account the given step, the next concepts of our ontology in criteria branch can be pairwise comparison and then pairwise comparison has pairwise comparison matrix. Next the eigenvectors and vectors of the local priority should be calculated (Table 2). Also the index of agreement should be found.

Table 2

Pairwise comparison matrix

	X1	X2	X3	X4	X5	X6	Eigenvectors	Normalized vectors of the local priority
X1	1	2	2	2	1	0.5	1.26	0.2
X2	0.5	1	1	1	0.5	0.333	0.66	0.1
X3	0.5	1	1	1	0.5	0.333	0.66	0.1
X4	0.5	1	1	1	0.5	0.333	0.66	0.1
X5	1	2	2	2	1	0.5	1.26	0.2
X6	2	3	3	3	2	1	2.18	0.3

Now it is necessary to obtain estimates of each alternative by each criterion. If you already have an objective assessment, then they are just issued, and normalized so that the amount is equal to 1. If the assessment is not objective, then, as the author (Saaty, 2011) writes, it is necessary to use pairwise comparison, similar to the previously discussed criteria. In this original task, an objective evaluation has already been given, which we also use. The following general priorities should be calculated. The result of this task is given in Table 3.

Table 3

The result of the task using AHP method

	X_1	X_2	X_3	X_4	X_5	X_6	Global priority
	Priority vector numerical value						
	0.2	0.1	0.1	0.1	0.2	0.3	
A_1	0.235	0.188	0.253	0.256	0.250	0.346	0.271
A_2	0.294	0.338	0.228	0.302	0.150	0.192	0.233
A_3	0.212	0.250	0.266	0.209	0.350	0.269	0.266
A_4	0.259	0.225	0.253	0.233	0.250	0.192	0.231

Considering this step we can define that alternative concept can have two branches: the first one is like criteria branch, the second one is selection of alternative, nodes evaluation and global priority. Using the previous discussed information, the ontology can be built as it is shown in Fig. 3.

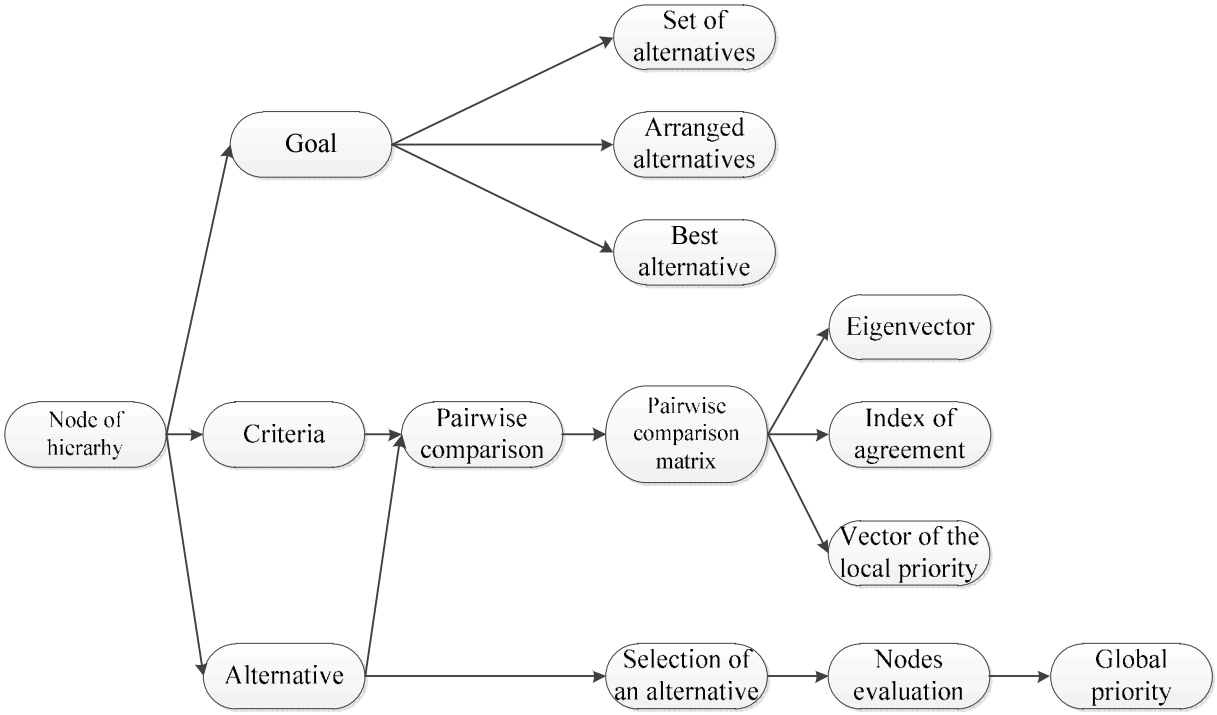


Fig.3. **AHP method ontology.**

Summary and future work

In this paper we have reviewed ontology engineering and ontology method. Based on the background study the method ontology has been built for AHP algorithm for decision support. The developed ontology can be reviewed and completed with some other elements. Our future research will include validation of the developed ontology using some domain ontology. Also we will work in a way to bridge the gap between XML and ontology and how to get from ontology to a specific XML schema.

Acknowledgements

Thanks to Dr.habil.sc.comp., Professor Arkady Borisov from Riga Technical University for help and support.

References

Aleksejeva, L.Y., 1998. Principles of Design and Procedures of Functioning of Decision Support Systems in Organizational Systems. Summary of PhD Thesis, Riga, 24 p.

Bolotova, L.S., 2012. Sistemy iskusstvennogo intellekta: Modeli i tehnologii, osnovannyye na znaniakh (Systems of Artificial Intelligence: Models and Technologies Based on Knowledge). Moscow, Finance and Statistics, 663 p. (in Russian)

Ferndndez, M. etc., 1997. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI Technical Report SS-97-06, pp. 33-40

Grechko, A.V., 2005. Ontologia metoda analiza iyerarhiy Saaty (Ontology of the Analytic Hierarchy Process of Saaty). Artificial intelligence 3, pp. 746 – 757. (in Russian)

Khakhalin, G.K., 2005. Predmetnaja ontologia dlya ponimania tekstov geometricheskih zadach (The Domain Ontology under Geometrical Text Understanding). Moscow, Research Centre of Electronic Computers, 9. p. Available at:
http://www.dialog-21.ru/digests/dialog2008/materials/html/Khakhalin.htm, 24.01.2013. (in Russian)

Kornyshova, Y., Deneckere, R., 2010. Decision-Making Ontology for Information Systems Engineering. ER'10 Proceedings of the 29th International Conference on Conceptual Modeling, pp. 104-107

Liu, O., Ma, J., 2005. A Web Services Approach to Model Management in DSS. *PACIS 2005 Proceedings, 31 p.*

Loranzo-Tello, A., Gomez-Perez, A., 2004. Ontometric: A Method to Choose the Appropriate Ontology. Journal of Database Management (JDM) Volume 15, Issue 2., p.18

Petrovsky, A.B., 2009. Gruppovoye mnogokriterialnoye prinyatiye resheniy s nesovpodayushchimi predpochteniami (Group Multiple Criteria Decision Making with Distinct Preferences). Scientific statements Belgorod State University 12, pp. 152-160. (in Russian)

Petrovskiy, A.B., 2009. Teorija prinyatiya resheniy (Decision Making Theory). Publishing Center “Academy”, 400 p. (in Russian)

Rothenfluh, T.E. etc, 1996. Reusable Ontologies, Knowledge-Acquisition Tools, and Performance Systems: PROTÉGÉ-II Solutions to Sisyphus-2. International Journal of Human-Computer Studies, Volume 44, Issue 3-4, pp. 303-332

Saaty, T., 2011. Prinyatie resheniy pri zavisemostyakh i obratnikh svyazyakh: Analiticheskiye seti. (Decision-Making in Dependence and Feedback: The Analytic Network). LIBROCOM, 360 p. (in Russian)

Uschold, M., King, M., 1995. Towards a Methodology for Building Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, 15 p.

Youn, S., McLeod, D., 2006. Ontology development tools for ontology-based knowledge management. Non-published research report, 100 p. Available at:
http://research.create.usc.edu/cgi/viewcontent.cgi?article=1104&context=nonpublished_reports, 24.01.2013

SEMI-Automatic approach to domain ontology building

Henrihs Gorskis, Tatjana Zmanovska, Jurijs Chizhov

Department of Modelling and Simulation, Riga Technical University, 1 Mezha street, Building 4, Room 456, Riga, LV-1048, Latvia
Henrihs.Gorskis@rtu.lv

Abstract: *This paper presents an automated method for building task ontology models from guideline models. A guideline is a specification of steps that need to be taken in certain situations and criteria that need to be fulfilled for these steps to be chosen. The ontology building process concentrates on machine readable guideline models, in particular, on the Guideline Interchange Format (GLIF). Since guidelines for similar processes most likely have many common concepts, it can be proposed that an ontological model of the task domain could be used for information storage and rule extraction.*
In order to accomplish the set goal of building a task ontology model from guidelines, it is necessary to do the following: create or convert basic concepts of the task into the concepts of ontology; create a relational structure between concepts within the ontology, capable of representing the choices and the sequence of the guidelines and unify equal concepts from different guidelines.
The extraction of concepts can be done by finding data request and task execution blocks in the guideline model. Data request blocks would correspond to environmental or system state concepts. To create a class hierarchy of the tasks and concepts for the ontology as well as the relational structure, a thorough analysis of the guidelines is required. This method creates a custom ontology structure and creates new relations that would be able to describe the processes in the guidelines in such a way that rule extraction is possible without keeping the original structure among guideline concepts.
Automated ontology building from guideline models seems to be very realistic and has the potential of combining the practical data from guidelines with the capabilities of ontology models, simultaneously uniting several guidelines into one large shared structure.

Keywords: Domain ontology building, clinical practice guideline.

Introduction

Ontology models are widely used in many scientific fields. Among many they are used in knowledge engineering, artificial intelligence, knowledge management, natural language processing, e-commerce, intelligent information integration, bio-informatics, education, semantic web etc. (Gorskis and Chizhov, 2012). The term ontology in computer science should not be confused with the philosophical meaning used by Plato and Aristotle, describing the nature of being. Building an ontology model is complex work and demands a lot of time. In order to build the ontology, usually a domain expert is required. The task of the expert is to declare all domain concepts and the relationships between them. The ontology is build for a certain field of interest and describes its domain by being a model representation of it. A computer readable ontology provides information for users and software agents about the domain. It can also be used to gain additional information about the domain by analyzing existing relationships between the given concepts.
A guideline is a document that dictates or helps determine a course of action by describing a situation and recommending the best action for that situation. To determine a situation, usually a set of criteria is given that need to be tested. There exist many computer readable document models and languages for the construction of guidelines. This paper proposes the use of these models and the guidelines themselves for the improvement and simplification, as well as acceleration of automated ontology building, by using them as prior information.
This paper takes a look at how computer understandable guidelines (and guidelines in general) can be used as the prior information and foundation of the ontology building process.

Guideline element models

There are several already developed guideline element models that are used to create guidelines. Most guidelines are used in the field of medicine and healthcare. Clinical guidelines are potential tools for standardizing patient care to improve its quality and cost effectiveness (Peleg et al., 2000). Structured, computer-interpretable guidelines can be delivered to the point of care in a way that enables decision support. This makes them very suitable for ontology building. Some of the most used guideline element models are GLIF, GLIF3 and GEM. The GLIF specification is aimed to provide a very precise representation of the guideline. It is designed to be both readable and computable. It is computable in the sense that the logic and sequence in guidelines specified in GLIF can not only be read, but also interpreted by computer. The very first version of GLIF was described in a pseudo code specifically meant to store values that describe the order of execution and related data. This high level code was almost ready for execution on a computer. Since the GLIF has moved on to a model that uses a

UML like language for structuring the guideline. This provides more flexibility, but has a far less clear data structure.
The GEM element model created by the SAGE project has defined an overall methodology for developing medical guideline knowledge bases (Shiffman et al., 2000). It is meant to provide decision support using these knowledge bases.
The GEM element model is meant to be used for guidelines specified in XML. It therefore can be depicted as a directed graph with the "Guideline Document" element as the root. The major concepts in the first tier of the GEM hierarchy below the root level are identity, developer, purpose, intended audience, method of development, target population, knowledge components, testing, and review plan. Each of these elements, in turn, comprises one or more additional levels of guideline constructs. Components of GEM are defined as XML elements. Elements have distinct names and are delimited with start and end tags.

Ontology engineering

An ontology is a specification of a conceptualization. That is, it is a description of concepts and relationships that exists for an agent or a community of agents (Gorskis and Chizhov, 2012). One can also say that the ontology is a formal, explicit description of concepts in a domain of discourse. It consists of classes sometimes called concepts; properties of each concept describing various features and attributes of the concept called slots, roles or properties; and restrictions on slots called facets or called role restrictions. The use, structure and functions of ontology models can be very different; however, the elements contained in ontology models are mostly the same. Among the many ontology definitions available, a clear description is in the following definition: an ontology structure is a 7-tuple O (1)

$$O = \{C, R, H^C, rel, I_C, I_R, A^O\} \quad (1)$$

that contains

- two disjoint sets C and R whose elements are called concepts and relations, respectively;
- a concept hierarchy H^C where H^C represents the hierarchy of concepts in their relations as $H^C \subseteq C \times C$, where H(C1, C2) means that C1 is a sub concept of C2. This hierarchy between concepts can also be called taxonomy;
- a function relation rel: R !C ×C, that relates concepts non-hierarchically. This can be written rel(R)=(C1,C2) or R(C1,C2) Relations that are not within the hierarchical structure and describe a relation between any 2 concepts or 2 individuals. These relations also include attributes;
- the sets I_C and I_R describe instances of the existing concepts (C) and relations (R);
- a set of ontology axioms A^O (additional rules and restrictions), expressed in an appropriate logical language.

From this it becomes clear that an ontology model holds concepts or classes that describe object or phenomena within a domain. The relations between the concepts inform about the hierarchical structure of classes and subclasses and also of other kinds of relations between them. This gives a description not only about the concepts and their place within the model and relative to other concepts, but also a description of the domain itself.
In order to represent ontology models, the use of a standardized language is needed. The Web Ontology Language (OWL) offers the necessary building blocks for ontology representation. The result of an automated or semi automated ontology building process can be an OWL file that holds the structured data of the ontology. OWL extends the Resource Description Framework (RDF) and is based on XML. That means that the ontology is stored as a text file in which the data are structured with the help of tags.

Related works

Even though the ontology building process described in this paper is using preexisting machine readable data, this work concerns all forms of ontology building. This process relies more on the analysis of the given data execution path and does not require data mining approaches, however it could be possible that in future such a necessity could arise.
By using guidelines as preexisting data for ontology building, the result of the process is a guideline stored within an ontology model. It conceptualizes the elements of a guideline, their properties, and defines the relationships that are held among them. For example, all medical guideline representation ontology models have a set of medical decisions and relevant actions (concepts), and a set of temporal rules that relate decision evaluation results to associated actions (relationships). A well established and generally acknowledged guideline representation ontology ensures that the resulting representations can be easily understood by non-authoring human readers, therefore facilitates the dissemination of guidelines across institutions. Well defined computational ontologies also provide considerable promise of enabling automated guideline acquisition,

visualization, execution, and sharing. Such characteristics are prerequisites for a computer-recognizable, interchangeable guideline format. Without these features, it is difficult to enable automated knowledge acquisition and execution for Clinical Decision Support Systems designed to enhance evidence-based practice (Zielstorff, 1998).

Only in the case where an element model for a given guideline is missing and the structure of the guideline itself is not very clear, a data mining approach could be necessary. In the paper by B. Fortuna the ontology building process uses information given by the classification system as hints for how to structure the ontology (Fortuna et al., 2006). The classification is based on SVM with respect to concepts of the domain. For such task the OntoGen software can be used. It is a semi-automatic and data-driven ontology editor focusing on the editing of topic ontologies. The system combines the text mining techniques with an efficient user interface to bridge the gap between the complex ontology editing tools and the domain experts who are constructing the ontology.

E. Blomqvist uses clues given by the concepts of a subject domain (Blomqvist, 2007). They are derived from the analysis of the subject areas with the help of case-based reasoning. Therefore, this approach is also semi automatic.

Suggested approach

The main idea of this paper is to use pre-existing and accessible guide element models as prior information in the ontology construction process. Using the predefined models that describe guideline elements for the basis of ontology building, it is possible to build an ontology model of the information given in the guideline, based on the underlying element structure.

First, it is necessary to obtain the guideline element model used for the creation of the guideline. There are different approaches for this. The simplest way is to obtain the original specification. This is easy if the guideline was written with the GEM or GLIF specification. The use of a predefined guideline element model can lead to excessive concepts in the final ontology in cases where a closely defined ontology would be desired.

When the predefined guideline element model is not given, it is required to create ontology concepts from the elements given in the guideline itself.

Using this approach the main structure of the ontology will be dictated by the element model. First, the main concept is defined. This concept can be a “thing”, or a general ontology concept. All other ontology concepts are related to this main concept. The other concepts are taken from the element model. Elements are translated into ontology concepts and the hierarchical structure of these concepts mimic the structure of the elements from the guideline element model. This first part has simplified the process of ontology building a lot, by virtue of being a translation and copying process.

Depending on the language of the guideline element model it is necessary to create transformation descriptions. For example, GEM is given as an XML schema. By using the information in the GEM file and looking for the element tag it is possible to extract the main concepts and relations. It is also possible to extract concept properties by looking for other information inside the element tag. The extraction of properties and the formats of properties can be more difficult than the extraction of elements alone.

This step needs to be supervised by an expert in order to make sure that only relevant information is being extracted from the model. The expert also needs to make sure that the relations between concepts are in order. In case of an ill defined model, the expert needs to create the necessary relations and add or remove information from the transformation.

The final step of ontology building from guidelines is the creation of concepts instances. For this it is required to be informed of all concepts within the guideline from the element model. By going through the guideline and reviewing the element found in it, instances of the related concept are created in the ontology and filled with the information provided in the guideline description.

Tasks to be solved

In order to extract the basic foundation of the ontology from GEM or any other specification a computerized solution must be created that transforms the GEM or other schema file into the ontology basis as an OWL file. A transformation needs to be created that finds <xs:element> tags and created <owl:class> tags that are given the same name as the “name” property in the guideline element. Further a solution for annotation and documentation needs to be created since it is unclear how there elements need to be treated in the ontology. In the case of GEM tags such as “<xs: complexType>” or “<xs:sequence>” can be overlooked since they are schema specific, but do not add information to the guideline model. All elements found within other elements can be labeled as subclasses. Since instances of elements in the guideline differ slightly from the GEM element model some slight changes can be required. The main difference is that connections between guideline element instances are not defined in the GEM element model. GLIF elements are defined in such a way that connections between elements are expected and regulated. A solution needs to be found for all cases. For a better use of the capabilities provided by the ontology it would be important to create conceptual relations for the guideline concepts as given in the guideline. For that, instances of concepts in the ontology can be viewed as concepts themselves. Relations between such concepts would be defined by the order of steps given in the guideline.

Since the order of steps in the guideline is defined by connections between actions and conditions, the information for rule creation can be extracted by following the execution of the guideline. Having extracted all concepts it is possible to detect what concepts are connected to the conditional actions and create firm or generalized rules based on those concepts. There rules are stored separately from the ontology in the IF/THEN syntax.
Since guideline element models can be very different both in their structure and the way the elements are described, it may be impossible to completely automate the extraction process. Automation at this time may be only capable to extract the general structure of the element model. User input can clarify the exact structure and meaning of the elements in the model.

Implementation of the suggested idea

In order to show the approach of ontology building from a guideline, a small test guideline is used. Fig. 1 shows a simplified guideline element model that will be used in this paper, which describes the overarching classes of information types and the elements that need to be filled with information for a guideline instance.
The first step of the ontology building process is to determine the guideline element model, since the base of the ontology will be based on it.

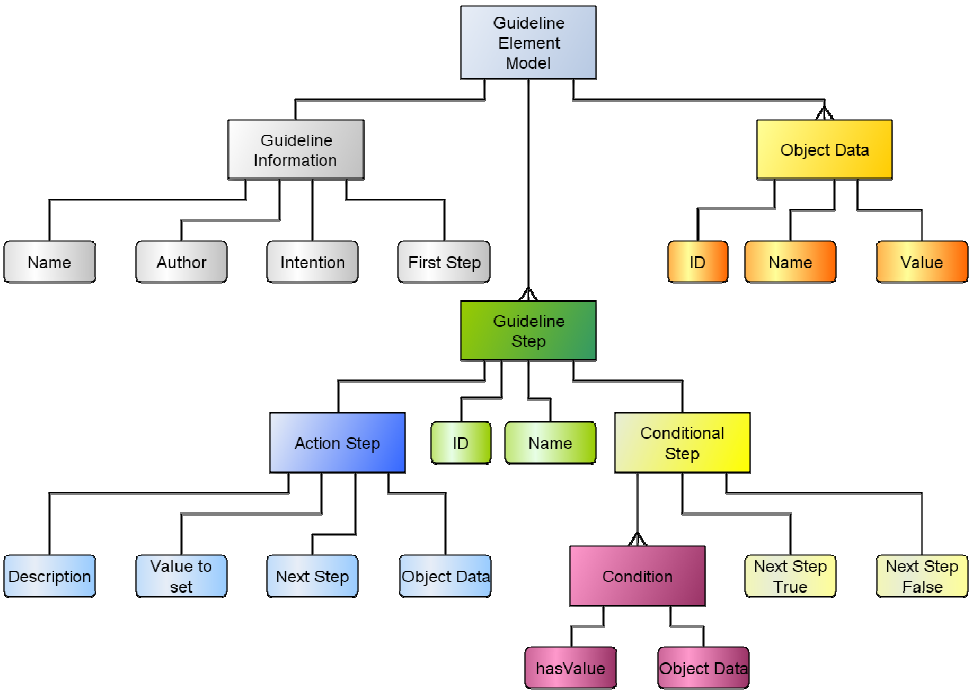


Fig. 1. **Test-guideline element model.**

This simplified element model was created based on a simplification of GEM, hoverer elements like “First step” and “Next step” are taken from GLIF. Every leaf element in this tree describes a certain value that has to be given. The element “Name” is part of the Guideline information, where Guideline information is a general class that does not have any separate information besides its leaf elements. It is important to note that the “name” element from the guidelines information is not the same as the “name” element from the guideline step or object data classes. The “next step” type elements hold information, which is required for the execution of a guideline written with this model. By analyzing the base structure of a guideline written using this element model or by analyzing the guideline itself, it is possible to recognize the root element, classes and subclasses, and inheritance between classes. Right away the ontology building process can start creating ontology classes by copying this information and adding the required notations. For example, the element <guideline.information> would be transformed to the OWL code in Fig. 2.

```
<owl:Class rdf:ID=" guideline.information ">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="guideline.element.model"/>
  </rdfs:subClassOf>
</owl:Class>
```

Fig. 2. **The OWL code example.**

The next step involves the guideline itself. Fig.3 shows a simple guideline about the weather. This is a simple sample guideline about the weather. It will be used to expand on the base structure of the ontology and provide the instance information for the existing concepts.

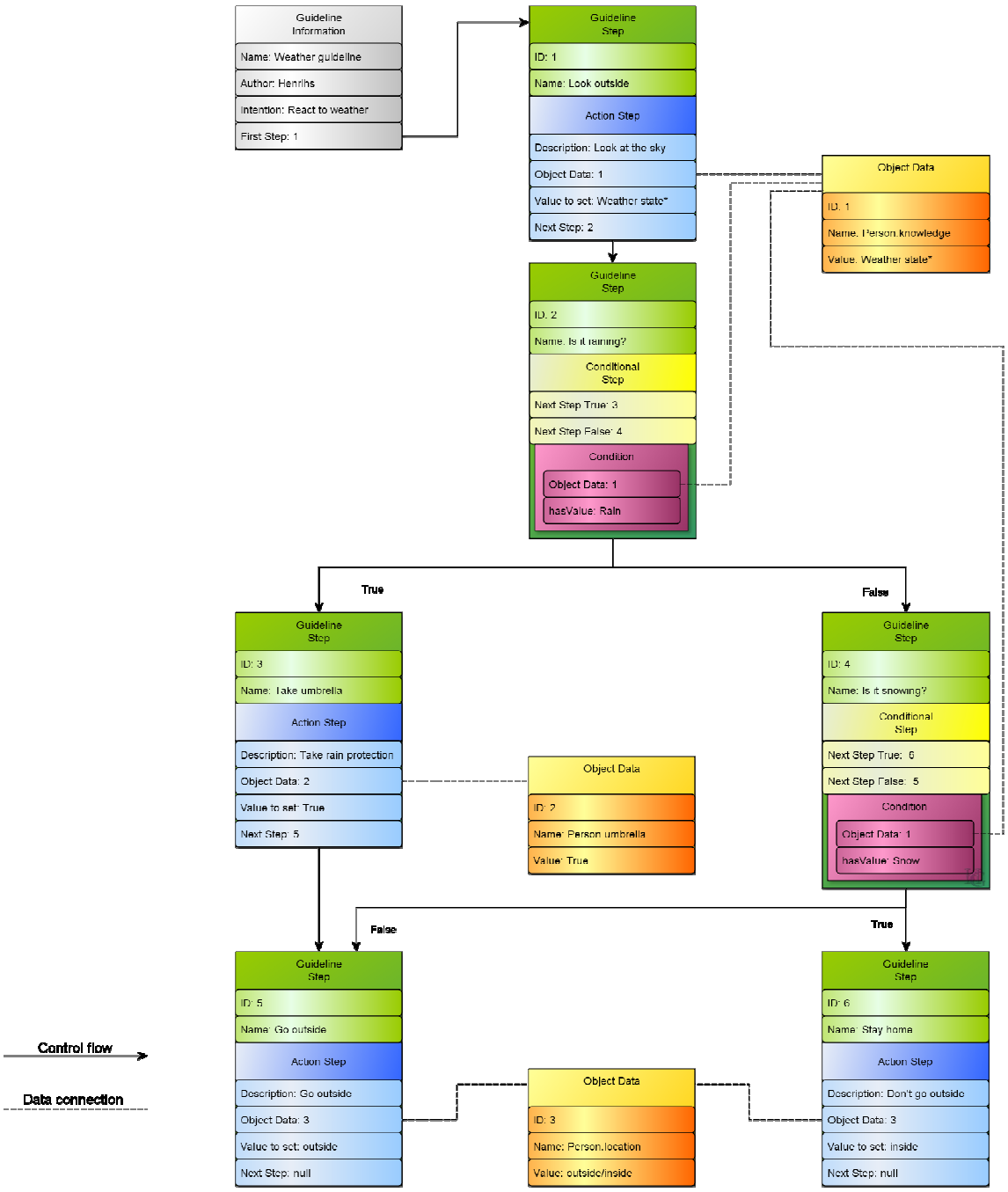


Fig. 3. **A test guideline.**

As one can see, the guideline shown in Fig. 3 consists of elements from the guideline element model. There is only one instance of the "guideline information" element and several instances of the "action step", "conditional step" and "object data" elements. The xml code of this guideline follows the same structure of the guideline element model. Any elements that are used more than one time are written separately and are filled with the information that is shown in Fig. 3. Since the "action step" and "conditional step" elements both inherit from the element "guideline step", the xml code for these elements will begin with the mention of the parent "guideline step" elements and hold information for "ID" and "name", however the body of the "guideline step" will only feature an "action step" or an "conditional step" element, but not both.

The third step takes the information from the guideline and adds instance data to the ontology that is based on the element model as is shown in Fig. 4.

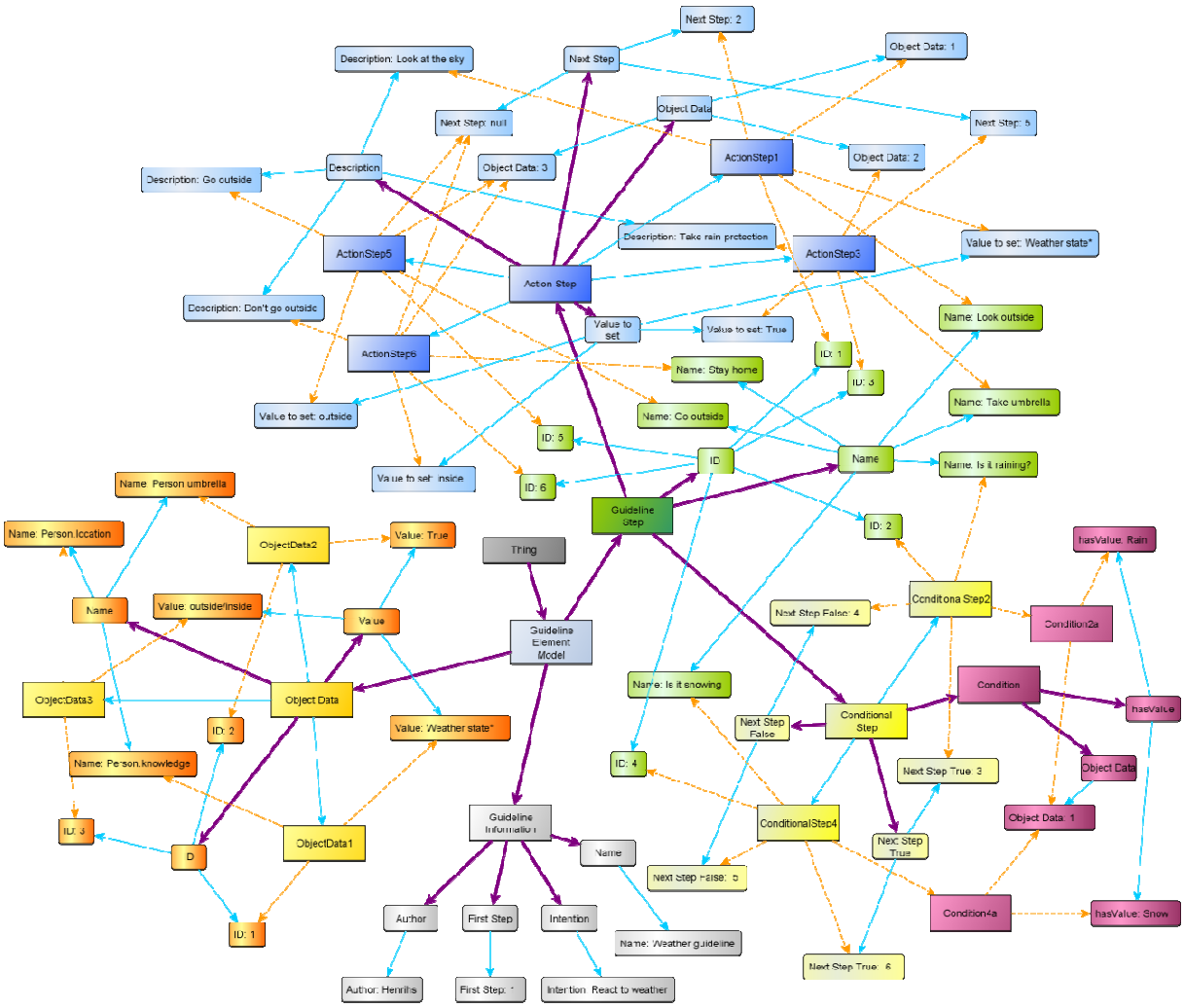


Fig. 4. **Test ontology from guideline data.**

All the information of the guideline has been implemented into ontology as instances of the data elements they belonged to. It should be noted that any classes that are similar in name, but are subclasses of different parent classes are not the same class. Any instance class that has the same information as an instance class that was already added is not includes, so that there is only one instance with the unique data. In order to maintain the execution information and the data described by the guideline additional classes have been created to store the relations between instances of data classes and instances of the overarching unifying classes. These are classes like “ActionStep1”, “ConditionalStep4” and “ObjectData3”. These are generated classes that are meant to store the relations between the abstract parent class and the instance of these classes in the guideline. These instances of the abstract parent classes are required to fulfill the idea of connecting instances only with other instances.

As the ontology is built, rule extraction operations can be performed in order to obtain additional rules about the guideline and the routes given in the mentioned guideline. As the guidelines data are analyzed, IF/THEN rules can be extracted about what data relate to each other in the context of the guidelines algorithm.

The resulting ontology should provide all the data that were given by the guideline and the element model it was based upon. This would conclude this method of ontology building.

Conclusion

This paper proposed a method of ontology construction by using information from and about a computer readable guideline. The obtained ontology summarizes the guideline elements and the information provided in the guideline itself. Even though ontology models like that could be usable right away, it is recommended and often even required to do additional work on the ontology by an expert.

The proposed method can be easily extended to use many guidelines of the same type for the acquisition of prior information. Any additional guideline would add more instance classes of the base classes. However in order to distinguish between information from different guidelines, additional information needs to be included in the ontology for the purpose of maintaining separation between data. One way of doing this is to add an instance of a “guideline” element that would be related to those and only those instances that were given in the separate guideline.

Future work can include the design and development of a computerized tool, which performs the described steps and generated the ontology model as OWL code.

Acknowledgements

Thanks to Dr.habil.sc.comp. Professor Arkady Borisov (Riga Technical University) for help and support.

References

Blomqvist, E., 2007. Semi-automatic Ontology Engineering using Patterns, ISWC/ASWC 2007 pp. 911-915.

Fortuna, B., Grobelnik, M., Mladenić, D., 2006. Semi-automatic data-driven ontology construction system, PASCAL EPRINTS – WORKING GROUP SUMMARY 15

Gorskis, H., Chizhov, Y., (2012). Ontology building using data mining techniques. Scientific Journal of Riga Technical University, Information Technology and Management Science, Vol.15, pp. 183-188.

Peleg, M., Boxwala, AA., Ogunyemi, O., 2000, GLIF3: the evolution of a guideline representation format. Proc AMIA Symp, pp. 645-649.

Shiffman, R., Karras, B., Agrawal, A., 2000. GEM: a proposal for a more comprehensive guideline document model using XML, J Am Med Inform Assoc., 7(5) pp. 488-98.

Syed Sibte Raza Abidi, Shapoor Shayegani, 2009. Modeling the Form and Function of Clinical Practice Guidelines: An Ontological Model to Computerize Clinical Practice Guidelines Knowledge Management for Health Care Procedures, Lecture Notes in Computer Science Volume 5626, pp. 81-91.

Zielstorff, R., 1998. Online practice guidelines: issues, obstacles, and future prospects, Journal of the American Medical Informatics Association, 5, pp. 227-236.

Using Fuzzy clustering with bioinformatics data

Madara Gasparovica[1], Ludmila Aleksejeva[1], Vladislavs Nazaruks[2]

[1] Department of Modelling and Simulation, Institute of Information Technology, Riga Technical University, Kalku str.1, Rīga, LV-1658, Latvia
[2]Department of Applied Computer Science, Riga Technical University, Kalku str.1, Rīga, LV-1658, Latvia
madara.gasparovica@rtu.lv, ludmila.aleksejeva_1@rtu.lv, vladislavs.nazaruks@rtu.lv

Abstract: *The article describes a research about fuzzy clustering algorithms, their creation and classification with the goal to determine the possibilities to use them in bioinformatics data clustering to find the membership of each record to a class. The study uses sixteen data sets used in previous studies by the authors and other researchers. Experiments were carried out using fuzzy c-means clustering method. The first section of the article gives an overview of the historical development of fuzzy clustering algorithms, their classification as well as the hypothesis that fuzzy clustering algorithms can be used to construct membership functions. The second section gives the description of the applied algorithm and the sixteen data sets used in the experiments. The third section gives a summary of the performed experiments and their results. And finally conclusions are drawn about the use of the algorithms in the clustering of bioinformatics data. The fourth section gives the overall conclusions and describes the further research directions. It is proven that fuzzy clustering algorithms (including the most popular – fuzzy c-means) can be used in membership function construction. Therefore fuzzy c-means algorithm with slight modifications can be used to construct membership functions of separate record attributes.*

Keywords: Fuzzy clustering, fuzzy c-means, bioinformatics data.

Introduction

Fuzzy logic is used in various scopes for problem solving where regular (crisp) algorithms cannot show good results. Often algorithms are modified adding fuzzy logic to improve their performance. For example, Prism algorithm (Cendrowska, 1987) was modified into a fuzzy version of itself and called FuzzyPrim (Wang et al., 1999), Bexa algorithm (Theron et al., 1996) has a fuzzy versions FuzzyBexaI (van Zyl et al., 2004a) and FuzzyBexaII (van Zyl et al., 2004 b). Also similar trend is in clustering – algorithm k-means (MacQueen, 1967) was modified into its fuzzy version Fuzzy c-means (Bezdek, 1981; Klir et al., 1997).
Clustering differs from classification in that most often there is no information about the membership of a record to a particular class. This information can be obtained by performing analysis to find groups in the data. Cluster analysis is grouping of objects into clusters in a way that similarity between two objects in one cluster is larger than similarity of two objects belonging to different clusters. The first fuzzy clustering method was created in 1974 (Dunn, 1974) and soon thereafter new methods were created and improved, which is also happening until now (e.g., Li et al., 2006; Gadaras et al., 2009; Lu et al., 2012).
This study explores fuzzy c-means (Bezdek, 1981) clustering algorithm. The article studies clustering algorithm main working principles and presents the applied experiments with 16 real bioinformatics data sets. The main goal of this study is to determine if the use of fuzzy classification algorithms (fuzzy c-means in particular) could be perspective in construction of membership functions. It also gives conclusions about clustering results and proves that the use of clustering can be perspective. It is also confirmed by (Bilgic et al., 1995), who included it into their research about membership function construction methods and concluding that fuzzy clustering should be applied to the output data and then projected onto input data, then clusters should be generated and the variable data of input-output relationships should be chosen. Then the membership functions are formed for the chosen variables and then the cross-validation is applied to the input data to evaluate the model. The author also venture a guess that most fuzzy clustering methods use Euclidean norm and the direction for future research would be to examine the use of other distance metrics (Bilgic et al., 1995).
For an overview research of various fuzzy clustering methods see (Ali et al., 2008), which also offers a classification of fuzzy clustering methods (Fig.1).
The second section of the article gives a description of the implemented fuzzy clustering methods and the data sets used in the experiments.

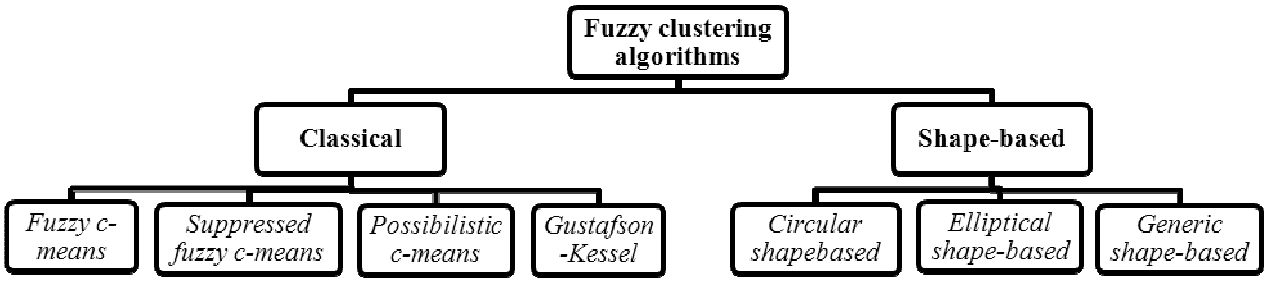


Fig. 1. **Fuzzy clustering algorithms; classification according to Ali.** (Ali et al., 2008)

The third section of this work gives a summary of the performed experiments and their results. It also gives conclusions about the use of algorithms in bioinformatics data clustering.
The fourth section gives overall conclusions and how the research can develop in the future studies.

Materials and methods

Fuzzy c-means clustering method

This study approaches practical analysis of clustering methods using bioinformatics data to evaluate the use of classical fuzzy clustering methods in the classification of bioinformatics data. Fuzzy c-means (fuzzy version of k-means) clustering was chosen as one of the top 10 data mining algorithms in the world (Wu et al., 2008). Therefore one can assume that Fuzzy c-means is one of the most used Top 10 algorithms.

Fuzzy c-means clustering algorithm tries to divide a finite element set $X = \{x_1, \dots, x_n\}$ into c cluster sets based on a particular criterion. If a finite data seti s given the algorithm returns a list of cluster canters $V = \{v_1; \dots; v_c\}$ and a conformity matrix $U = \{u_i(x) \| i = 1 \dots c; x \in X\}$, where each element $u_i(x)$ determines level at which an element x belongs to cluster c_i (Klir et al., 1995; Bezdek, 1981). Usually for each data element its sum of membership levels to different clusters equals one, meaning that it is normalized:

$$\forall x \left(\sum_i u_i(x) = 1 \right)$$

The centroid v_k of each cluster k is calculated as follows:

$$v_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}$$

When centroids of all clusters have been calculated the coefficients $u_k(x)$ that show the membership levels of a point for each point x can be calculated according to the following formula:

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(v_k; x)}{d(v_j; x)} \right)^{2/m-1}}$$

where $d(a; b)$ is the distance between data points a and b. Often the Euclidean distance is chosen for this:

$$d(a; b) \equiv \sqrt{\sum_{i=1}^{dim\, a} (a_i - b_i)^2}$$

m is the parameter of the real number that describes the influence on membership levels (Klir et al., 1995).
Algorithm fuzzy *c*-means itself is very similar to the rough *k*-means algorithm and can be described as follows:

1. Choose the number of clusters c.
2. Randomly assign membership coefficients $u_i(x)$ for each data point.
3. Repeat until the necessary accuracy is achieved:
    a. Calculate centroids v_i of all clusters.
    b. For each point calculate its membership coefficients $u_i(x)$.

To determine if the algorithm stops because the necessary accuracy ε is achieved, in the third step: calculate the absolute values $|\Delta u_k(x)|$ of changes of all coefficients $u_k(x)$ in the last iteration; find the maximum $\max_{k,x} |\Delta u_k(x)|$ of the absolute values and then compare to the accuracy ε that was chosen in the beginning of

the algorithm – if the maximum is not larger than ε, then the desired accuracy has been reached and the algorithm stops (Klir et al., 1995; Bezdek, 1981).

Used data sets

The study uses 16 popular data sets (Gasparovica et al,. 2012a; Gasparovica et al., 2012b) mentioned in literature (see Table 1): DLBCL data set (Shipp et al., 2002) - diffuse large B-cell lymphomas (DLBCL) and follicular lymphomas (FL); GSE2191 - acute myeloid leukemia prognosis after treatment (Yagi et al., 2003); GSE349_350 - breast cancer treatment response (Chang et al., 2003); GSE3726 - breast & colon cancer gene expressions (Chowdary et al., 2006); GSE89 - bladder cancer gene expressions, (Dyrskjøt et al., 2003). GSE2535 - chronic myeloid leukemia treatment response (Crossman et al., 2005); GSE967- childhood tumors - Ewing's sarcoma (EWS), embryonal and alveolar rhabdomyosarcoma (eRMS and aRMS) gene expression (Baer et al., 2004); GSE2685 - diffuse and intestinal gastric cancer gene expressions (Hippo et al., 2002); GSE1577 - lymphoma & leukemia (T-ALL, T-LL and B-ALL) gene expressions (Raetz et al., 2006); GSE1987 - lung cancer (Dehan et al., 2007).

Table 1

Used data sets

Name	Diagnostic classes	Number of genes	Genes after FCBFS	Number of samples
DLBCL	Diffuse large B-cell lymphoma (DLBCL): 58 ex.	7070	74	77
	Follicular lymphoma (FL): 19 ex.			
GSE2191 (AML prognosis)	Remission (remission): 28 ex.	12625	54	54
	Relapse (relapse): 26 ex.			
GSE349_350 (Breast cancer)	Resistant to docetaxel treatment (resistant): 14 ex.	12625	2	24
	Sensitive to docetaxel treatment (sensitive): 10 ex.			
GSE3726 (breast & colon cancer)	Breast cancer (breast): 31 ex.	22283	37	52
	Colon cancer (colon): 21 ex.			
GSE89 (bladder cancer)	Tumor stage T2-T4 (T2+): 10 ex.	5724	48	40
	Tumor stage Ta (Ta): 19 ex.			
	Tumor stage T1 (T1): 11 ex.			
GSE2535 (CML treatment)	Non-responder to imatinib treatment (Non-Responder): 12 ex.	12625	33	28
	Responder to imatinib treatment (Responder): 16 ex.			
GSE967 (childhood tumors)	Ewing's sarcoma (EWS): 11 ex.	9945	2	23
	Rhabdomyosarcoma (RMS): 12 ex.			
GSE967 (childhood tumors)	Ewing's sarcoma (EWS): 11 ex.	9945	28	23
	Embryonal rhabdomyosarcoma (eRMS): 3 ex.			
	Alveolar rhabdomyosarcoma (aRMS): 9 ex.			
GSE2685 (gastric cancer)	Normal gastric tissue (Normal): 8 ex.	4522	2	30
	Diffuse gastric tumor (Diffuse): 5 ex.			
	Intestinal gastric tumor (Intestinal): 17 ex.			
GSE2685 (gastric cancer)	Normal gastric tissue (Normal): 8 ex.	4522	25	30
	Advanced gastric cancer tissue (Tumor): 22 ex.			
GSE1577 (lymphoma & leukemia)	T-cell lymphoblastic lymphoma (T-LL): 9 ex.	15434	2	19
	T-cell acute lymphoblastic leukemia (T-ALL): 10 ex.			
GSE1577 (lymphoma & leukemia)	T-cell lymphoblastic lymphoma (T-LL): 9 ex.	15434	53	29
	T-cell acute lymphoblastic leukemia (T-ALL): 10			
	B-cell acute lymphoblastic leukemia (B-ALL): 10			
GSE1987 (lung cancer)	Squamous cell carcinoma (Squamous): 17 ex.	10541	41	34
	Adenocarcinoma (Adenocarcinoma): 8 ex.			
	Normal lung tissue (Normal): 9 ex.			

Name	Diagnostic classes	Number of genes	Genes after FCBFS	Number of samples
GSE468 (medulloblastoma)	Metastatic medulloblastoma (Met): 10 ex.	1465	20	23
	Non-metastatic medulloblastoma (NonMet): 13 ex.			
GSE2443 (prostate cancer)	Androgen dependent tumor (dependent): 10 ex.	12627	2	20
	Androgen - independent tumor (independent): 10			
Brain tumor	Medulloblastoma (medulloblastoma): 10 ex.	7129	67	40
	Malignant glioma (glioma): 10 ex.			
	Rhabdoid tumor (RhabdoidTu): 10 ex.			
	Normal cerebellum (Normal): 4 ex.			
	Primitive neuroectodermal tumor (PNET): 6 ex.			

GSE468 - metastatic (Met) or non-metastatic (NonMet) medulloblastoma gene expressions (MacDonald et al., 2001); GSE2443 - prostate cancer gene expressions (Best et al., 2005); Braintumor (Pomeroy et al., 2002) - distinguishes between different embryonal tumors of the central nervous system on the basis of DNA expression signatures. The popular data sets are provided by University of Ljubljana, Faculty of Computer and Information Science Bioinformatics Laboratory on their home page (University of Ljubljana, 2012). All popular data sets characterize gene expression data. The column 'Genes after FCBFS' shows the number of genes that were used in the second series of experiments after applying attribute selection method Fast correlation based filter solution (Yu at al., 2003; Gasparovica 2012a) to the initial data set.

Fuzzy clustering results and discussion

Difference between clustering and real data

The experiments were carried out in two sessions – the first used data sets reduced with Fast-correlation baser filter solution attribute selection method (Gasparovica, et al. 2012a), the second used full data sets shown in the table. Fuzzy c-means experiments were carried out with the following parameters: number of clusters from 2 to 10, ε=0.0001 for reduced sets (with FCBFS), and ε=0.001 for the full data sets.

Table 2

Difference between the original and the clustering prognosis

Name	Number of classes	Original data set	Data set with FCBFS
DLBCL	2	0.69	0.57
GSE2191 (AML prognosis)	2	0.59	0.57
GSE349_350 (Breast cancer)	2	0.83	0.79
GSE3726 (breast & colon cancer)	2	0.65	0.65
GSE89 (bladder cancer)	3	0.60	0.83
GSE2535 (CML treatment)	2	0.54	0.79
GSE967 (childhood tumors – 2cl.)	2	0.70	0.70
GSE967 (childhood tumors – 3cl.)	3	0.52	0.48
GSE2685 (gastric cancer -3cl.)	3	*	0.67
GSE2685 (gastric cancer – 2cl.)	2	0.93	0.90
GSE1577 (lymphoma & leukemia – 2cl.)	2	0.84	0.89
GSE1577 (lymphoma & leukemia – 3cl.)	3	0.72	0.97
GSE1987 (lung cancer)	3	**	0.85
GSE468 (medulloblastoma)	2	0.42	0.74
GSE2443 (prostate cancer)	2	0.70	0.75
Brain tumor	5	**	0.73

(*) The healthy class is well separated but both cancer classes have the same membership; (**) All classes have the same membership.

Table 2 shows the difference between original class and prognosis for each data set, respectively it shows how much the defined membership of each record to a specific class corresponds to the real class. Analysis of clustering results and comparison of C-means algorithm applied to real data sets and to FCBFS data sets, it can be seen that better results are acquired in the data sets reduced using FCBFS. It can be explained by the fact that the original data set holds many attributes that require additional computations and consideration of so many criteria that they are hard to balance.

Clustering results

To evaluate the clustering results obtained by the fuzzy *c*-means algorithm, the internal evaluation metric S, which is described further, is used.
For degrees of belonging $u_k(x)$ of each data item x, the standard deviation σ is calculated (Weisstein):

$$\sigma(x) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(u_k(x) - \overline{u_k(x)}\right)^2}$$

where K is a number of clusters. As the values of $u_k(x)$ are normalized, the arithmetic mean $\overline{u_k(x)}$ equals to $\frac{1}{K}$; therefore, $\sigma(x) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(u_k(x) - \frac{1}{K}\right)^2}$. The sense of the metric σ is the following: for a data item x it equals to 0 if (and only if) the degrees of belonging of this data item to all clusters are equal ($u_1(x) = \cdots = u_K(x)$); and it takes the maximal value, if the data item belongs to one cluster with the degree of 1: $u_l(x) = 1$, $\forall k \neq l(u_k(x) = 0)$.
To compare the values of σ of the same data items independently of different numbers of clusters K, σ needs to be normalized in order to take values in the interval $[0; 1]$. The maximal value of σ is $\sigma_{max} = \sqrt{\frac{1}{K}\left(\left(1 - \frac{1}{K}\right)^2 + (K-1)\left(0 - \frac{1}{K}\right)^2\right)} = \frac{\sqrt{K-1}}{K}$. Therefore, the normalized value of σ is $\sigma_n = \frac{\sigma(x)}{\sigma_{max}}$.
For all data items x, the metric S is defined as the arithmetic average of $\sigma(x)$: $S = \overline{\sigma(x)}$. It takes the value of 0 (or near it), if the clustering is poor (the data item is not associated with any cluster with a significant certainty), and the value of 1 (or near it), if the clustering is authoritative (however, not obligatory true); the higher value of the metric is the better one.
The results of applying the metric S to the clustering results obtained by the fuzzy *c*-means algorithm are showed on the following two figures (two figures were used in order to improve the appearance of the graphs):

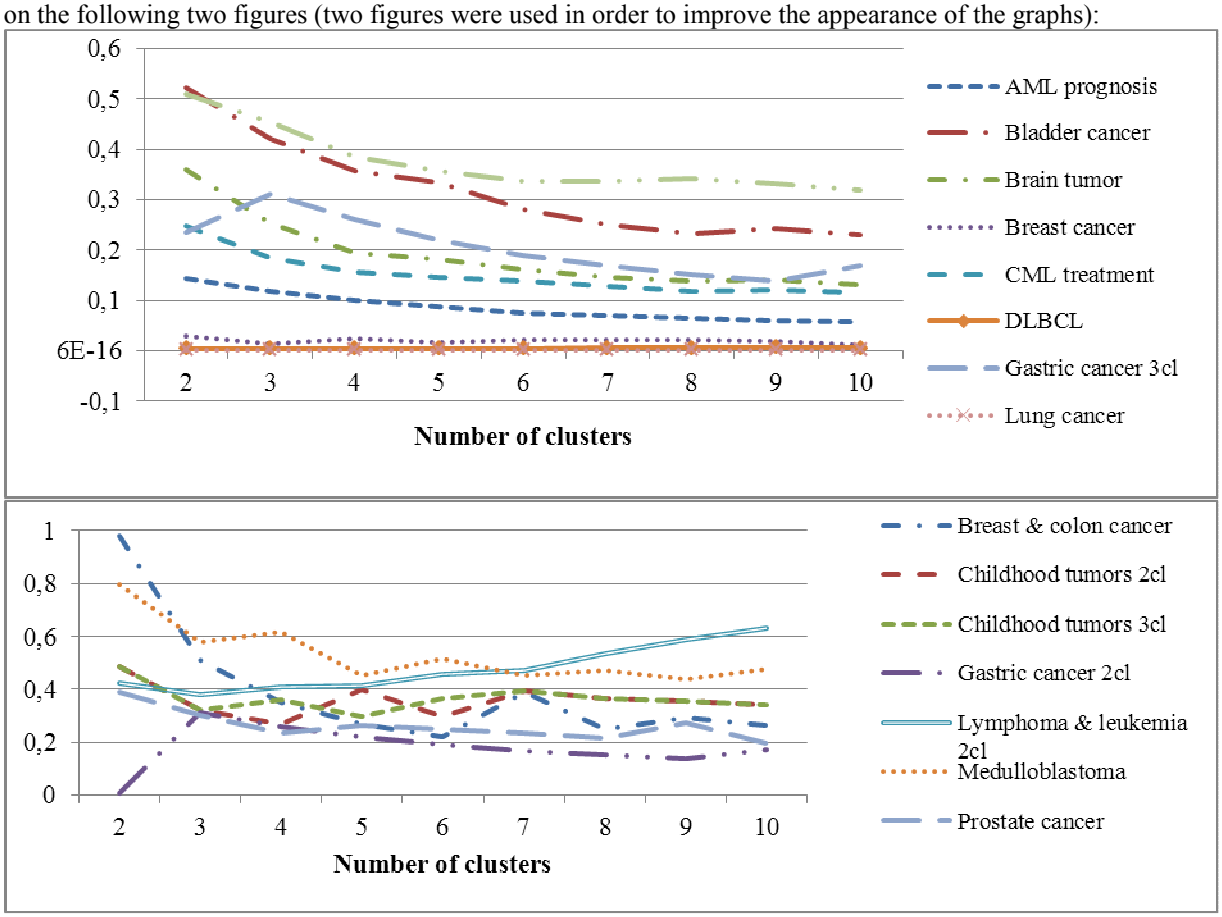


Fig. 2. **Fuzzy c-means results of applying the metric S.**

As one can see, the most common tendency of S depending on the number of clusters is to decrease. (The exception is the *Lymphoma leukemia 2cl* data set; however, it can be explained with a low (19) number of data items in the data set.) The maximums on the graphs show the numbers of clusters where the fuzzy *c*-means

algorithm can potentially give best results. For example, for the majority of data sets, the potentially best number of clusters is 2 (except for the *Gastric cancer 3cl* and *Gastric cancer 2cl* data sets).
Speaking about the absolute values of the metric, one can conclude that for the *Lung cancer*, *Breast cancer* and *CML treatment* data sets the fuzzy *c*-means clustering algorithm cannot give good results; however, for the *Breast & colon cancer*, *Medulloblastoma*, *Bladder cancer*, *Lymphoma leukemia 3cl* and *Childhood tumor 3kl* data sets this algorithm can potentially give good results. (The word 'potentially' here shows that we used the internal evaluation (not the external one), where known cluster labels were not compared with the clustering results.)

Iteration and Cluster number experiments

While conducting the experiments it was decided to also study the interactions of number of iterations and the number of clusters. As it can be seen in Fig.3 precise conclusions cannot be drawn but the data set behaviours have at least two trends. As it is shown in the upper graph in Fig. 3, the numbers of iterations display sharp fluctuations when the number of clusters is increased and then the number of iterations returns to the initial state, and beginning from three clusters the number of iterations for each number of clusters grows. This lets concluding that there is a number of clusters when the number of iterations increases very sharply. The lower graph in Fig. 3 shows a different trend – there are several numbers of clusters that have high numbers of iterations.

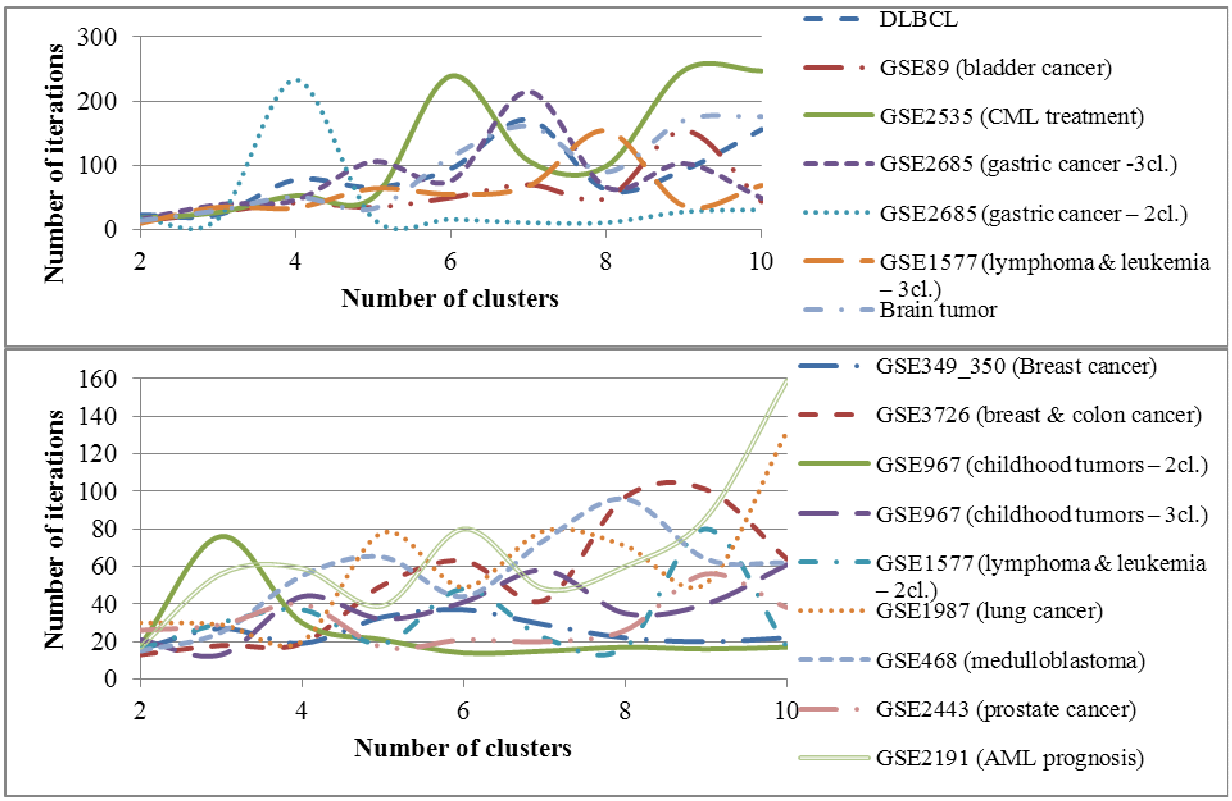


Fig. 3. **Iteration and cluster number experiments data sets with FCBFS.**

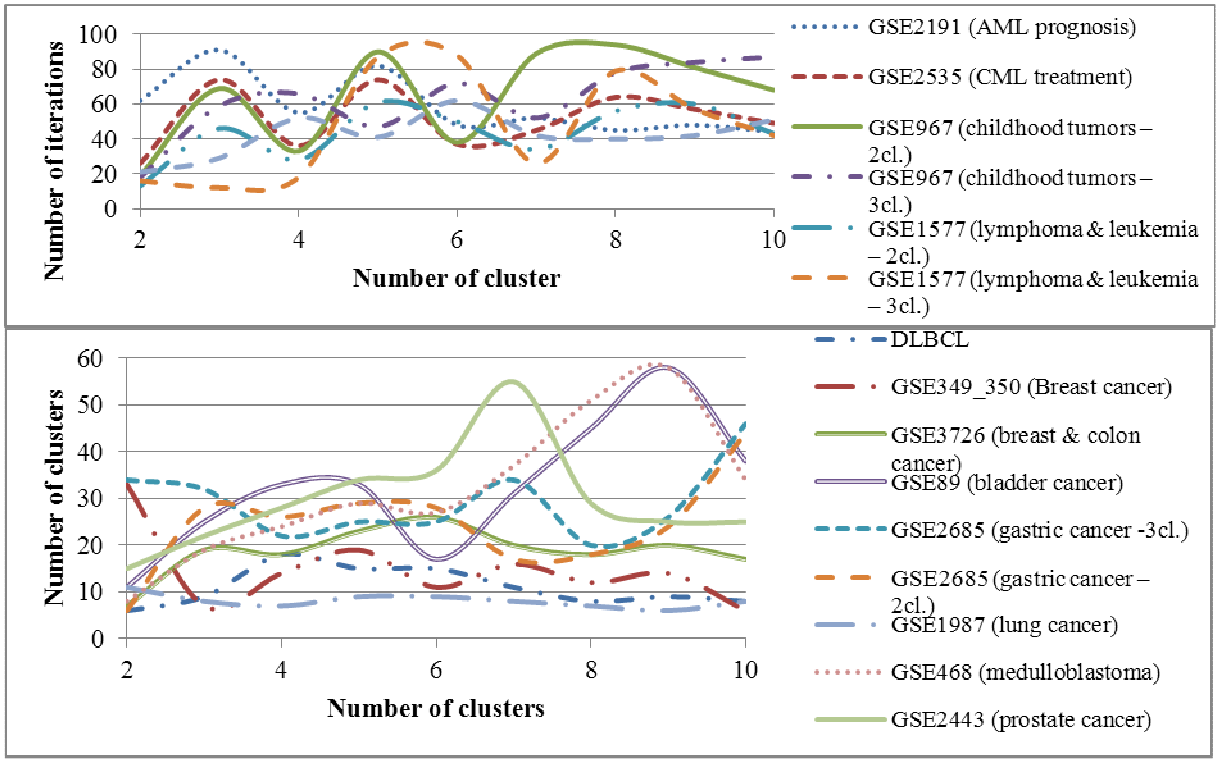


Fig. 4. **Iteration and cluster number experiments original data sets.**

If we look at graphs shown in Fig.4 that depict the trends in the original data sets with a slightly lower accuracy, it can be seen that the number of iterations to reach this accuracy is smaller. But one can see analogical trends in these data – some data sets have one number of clusters that requires much more iterations, but other show fluctuations in a small interval.

Conclusion

The use of fuzzy clustering algorithms while determining class membership is perspective because the acquired membership functions can be used in further research. A question that should be addressed in future is how to transfer information about the class weights in the process of constructing membership functions for the initial training data.

The research about the relations between the number of clusters and the number of iterations shows that there are certain relations but strong conclusions cannot be drawn. To do that there should be additional experimental research carried out.

Comparison of full and reduced data set clustering results shows that in future studies it is more perspective to use data sets reduced with FCBFS because the comparative parameters are better than in the full data sets.

The research about other membership function construction methods and algorithms should be continued in order to improve the shortcomings of fuzzy c- means algorithm.

Acknowledgements

Thanks to Dr.habil.sc.comp. Professor Arkady Borisov (Riga Technical University) for help and support.

References

Ali, A., Karmakar M., Gour C., Dooley, L.S., 2008. Review on Fuzzy Clustering Algorithms. *Journal of Advanced Computations*, Vol.2, No.3, pp. 169–181.

Baer, C., Nees, M., Breit, S., Selle, B. et al., 2004 Jul 10. Profiling and functional annotation of mRNA gene expression in pediatric rhabdomyosarcoma and Ewing's sarcoma. Int J Cancer;110(5), pp.687-694.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers Norwell, MA, USA, p. 256.

Bilgic, T., Türksen, I.B., 2005. Measurement Of Membership Functions: Theoretical And Empirical Work. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.6691, Accessed 05.01.2013.

Cendrowska J., 1987, PRISM: An Algorithm for Inducing Modular Rules. *International Journal of Human-computer Studies / International Journal of Man-machine Studies - IJMMS*, Vol. 27, no. 4, pp. 349-370.

Chang, JC., Wooten, EC., Tsimelzon, A., Hilsenbeck, SG. et al., 2003 Aug 2. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet. 362(9381), pp.362 -369.

Chowdary, D., Lathrop, J., Skelton, J., Curtin, K. et al., 2006 Feb, Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative. J Mol Diagn;8(1), pp.31-39.

Crossman, L.C., Mori, M., Hsieh, Y.C., Lange, T. et al., 2005 Apr. In chronic myeloid leukemia white cells from cytogenetic responders and non-responders to imatinib have very similar gene expression signatures. Haematologica;90(4), pp.459-464.

Dehan, E., Ben-Dor, A., Liao, W., Lipson, D. et al., 2007 May. Chromosomal aberrations and gene expression profiles in non-small cell lung cancer. Lung Cancer;56(2), pp.175-184

Dunn, J.C., 1974. Well-Separated Clusters and Optimal Fuzzy Partitions, *Journal of Cybernetics and Systems*, vol. 4, pp.95-104.

Dyrskjøt L., Thykjaer T., Kruhøffer M., Jensen JL., et al., 2003 Jan, Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet;33(1), pp.90-96.

Gadaras, I., Mikhailov, L. 2009. An interpretable fuzzy rule-based classification methodology for medical diagnosis. *Artificial Intelligence in Medicine*, Vol. 47, pp. 25 – 41.

Gasparovica, M., Aleksejeva, L., 2012. Feature Selection for Bioinformatics Data Sets – Is It Recommended? *Proceedings of the 5th International Conference on Applied Information and Communication Technologies (AICT2012)*, Latvia, Jelgava, pp. 325-335.

Gasparovica, M., Aleksejeva, L., Gersons, V. 2012. The Use of BEXA Family Algorithms in Bioinformatics Data Classification. *Scientific Journal of RTU*. 5. series., Information Technology and Management Science, vol. 15, pp 120-126.

Hippo, Y., Taniguchi, H., Tsutsumi, S., Machida, N. et al., 2002 Jan 1. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. Cancer Res;62(1), pp.233-240.

http://www.biolab.si/supp/bi-cancer/projections/index.htm, Accessed: 02.01.2012.

Klir, J.G., Yuan B., 1997. Fuzzy Sets and Fuzzy Logis. Theory and Applications. Prentice Hall of India. Private Limited, p. 592.

Klir, J.G., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logis. Theory and Applications. Prentice Hall PTR, Upper Saddle River, NJ, p. 592.

Li, X.L., Tan, Y.C., Ng, S.K., 2006. Systematic Gene Function Prediction Using a Fuzzy Nearest-Cluster Method on Gene Expression Data. Proceedings of the First International Multi- Symposiums on Computer and Computational Sciences (IMSCCS|06), pp.171 - 178

Lu, W., Rankin, J.G., Bondra, A., Trader, C., Heeren, A., Harrington, P. de B. 2012. Ignitable liquid identification using gas chromatography/mass spectrometry data by projected difference resolution mapping and fuzzy rule-building expert system classification. Forensic science international, Vol.220, pp. 210-218.

MacDonald, T.J., Brown, K.M., LaFleur, B., Peterson, K. et al., 2001 Oct. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. Nat Genet;29(2), pp.143-152.

MacQueen, J. 1967. Some methods for classification and analysis ofvmultivariate observations. *Proceedings of the Fifth Berkeley Symp. On Math. Stat. and Prob.*, 1, pp. 281-296.

Pomeroy, S.L, Tamayo, P., Gaasenbeek, M., et al., 2002 Jan 24. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature.;415(6870), pp.436-442.

Raetz, E.A., Perkins, S.L., Bhojwani, D., Smock, K. et al., 2006 Aug. Gene expression profiling reveals intrinsic differences between T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma. Pediatr Blood Cancer;47(2), pp.130-40.

Shipp, M.A., et al., January 2002. Diffuse Large B-Cell Lymphoma Outcome Prediction by Gene-Expression Profiling and Supervised Machine Learning. Nat Med., Vol.8, No.1, pp. 68-74.

Theron, H., Cloete, I., 1996. BEXA: A covering algorithm for learning propositional concept descriptions. *Machine Learning*, Vol. 24, No. 1, pp 5-40.

University of Ljubljana, Faculty of computer and information science Bioinformatics Laboratory home page:

van Zyl, J., Cloete, I., 2004. FuzzConRI - A Fuzzy Conjunctive Rule Inducer. *Proceedings of the ECML/PKDD-04 Workshop on Advances in Inductive Rule Learning*, pp.194-203.

van Zyl, J., Cloete, I., 2004. Simultaneous Concept Learning of Fuzzy Rules. *Proceedings of the Fifteenth European Conference on Machine Learning,* pp. 548-559.

Wang, C. H., Liu, J. F., Hong, T. P., Tseng, S.S., 1999. A fuzzy inductive learning strategy for modular rules, *Fuzzy sets and Systems*, 103, pp. 91–105.

Wu, X.,·Kumar, V.,·Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, S.H., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. *Knowl Inf Syst* ,14, pp.1–37.

Weisstein, E. W. Standard Deviation. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/StandardDeviation.html, Accessed: 05.01.2013.

Yagi, T., Morimoto, A., Eguchi, M., Hibi, S. et al., 2003 Sep 1. Identification of a gene expression signature associated with pediatric AML prognosis. Blood;102(5), pp.1849-1856.

Yu, L., and Liu, H., 2003. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), August 21-24, 2003, Washington DC. AAAI Press, Menlo Park, California.

.

MODELLING AND SIMULATION TECHNOLOGIES

Adaptability of attractor selection in virtual network topology control method

Dmitrijs Finaskins

Telecommunications institute, Riga Technical University, Azenes street 12, Riga, LV-1048, Latvia
Dmitrijs.finaskins@rtu.lv

Abstract: *Telecommunication network topologies are dynamically changing nowadays. The core telecommunication networks are based on fiber optics infrastructure. WDM technology is used to transmit multiple flows of traffic over a single fiber. Virtual network topology (VNT) is used to route IP traffic over WDM networks. VNT control system must be adaptive to traffic changes and dynamically reconfigure VNT. This can be achieved by using attractor selection, which models biological system behavior. Biological systems are extremely adaptive to unknown changes in environmental conditions. This adaptability could be used as a model for VNT control system. Such VNT control method is based on a system of stochastic differential equations, which use load on links, retrieved by SNMP, and stochastic fluctuations – Gaussian noise – as input parameters to dynamically reconfigure VNT and optimize its performance. Network simulations show that this VNT control method is very quick and adaptable to traffic changes.*

Keywords: Virtual Network Topology, VNT, Adaptability, Attractor, Stochastic Differential Equation, SDE.

Introduction

Core telecommunication networks nowadays are based on Wavelength Division Multiplexing (WDM) technology. In such division multiplexing technology transmitting/receiving channels are divided by wave length. It allows transmitting multiple traffic channels over a single fiber. Traffic by fiber optics can be transmitted over long distances without any additional equipment. Because of that WDM networks are commonly used to carry Internet traffic at backbone level. The major protocol of Internet is IP protocol. One of possibilities to carry IP traffic over WDM network is to construct virtual network topology, which includes transmitting/receiving channels (lightpaths) and IP routers. There are many VNT control methods, which configure/reconfigure VNT according to traffic demand matrices. Traffic demand matrices show how traffic flows are distributed via lightpaths.

One of the most exciting Internet opportunities for end-users is to share their pictures, videos and so on with other users; to communicate with each other using online services such as Skype. These cause constant and rapid changes in traffic flows between IP routers in VNT. There is a need to reconfigure VNT over a period of time in order to provide high-level service with minimal delays. VNT needs to be adaptable to changes in traffic demand.

There are two modes of constructing traffic demand matrices – offline and online or dynamical mode. In offline approach traffic demand matrices are constructed using previous known information about changes in traffic demand (Agrawal et al., 2006; Chen et al., 2008; Ricciato et al., 2002). The major weakness of this approach – offline methods will not work correctly if traffic flow changes will be different from ones expected. Online approach allows reconfiguring VNT dynamically. In this case periodical measurement results are used. To evaluate VNT status, information about average or maximum link utilization, packet delays can be used. Based on this information new lightpaths are added to source-destination pair of nodes if, for example, link utilization between these nodes is more than threshold and deleted if lightpath is underutilized.

The majority of online VNT control methods are used when traffic demand is changing periodically and gradually (Lakhina et al., 2004). This approach will not work if changes of traffic demand are not predictable. There is a need to develop such VNT control method, which is adaptable to unknown changes in network environment.

One of the approaches is to use attractor selection, which represents mechanism of adaptation to unknown changes of biological systems (Koizumi et al., 2008). The main idea of attractor selection – the system is driven by two components – deterministic and stochastic. Attractors are a part of the equilibrium points in the solution space. Conditions of such system are controlled by very simple feedback. When conditions of a system are suitable (close to one of the attractors), it is driven almost only by deterministic behavior, stochastic influence is very limited. When conditions of the systems are poor, deterministic behavior influence is close to zero and in this case system is driven by stochastic behavior. It randomly fluctuates searching for a new attractor. When this attractor is found, deterministic behavior again dominates over stochastic.

In this paper a research of attractor selection method is presented, which is adopted from (Furusawa et al., 2008) and described in (Koizumi et al., 2008). The object of the paper - how adaptability of VNT, controlled by attractor selection, depends on variety of parameters. The paper is organized as follows. In the next section

attractor selection mechanism is briefly described and the rest of the paper is devoted to network simulations and result analysis.

Attractor selection mechanism

Every pair of nodes, between which connection can be established, is represented by control unit u_{ij}, where i and j are indexes of nodes. Every control unit has its control value $x_{u_{ij}}$ Indexes s and d refers to source and destination nodes.

How control values change over time can be expressed by differential equation, also known as Langevin equation (Koizumi et al., 2008):

$$\frac{dx_{u_{ij}}}{dt} = v_g \cdot f\left(\sum_{u_{sd}} W(u_{ij}, u_{sd}) \cdot x_{u_{sd}} - \theta_{u_{ij}}\right) - v_g \cdot x_{u_{ij}} + \eta \quad (1)$$

where v_g – indicates conditions of VNT ,
$W(u_{ij}, u_{sd})$ – regulatory matrix of VNT,
$\theta_{u_{ij}}$ – coefficient, which depends on minimum and maximum loads on links in VNT,
η – Gaussian noise.

As it was used in (Koizumi et al., 2008), Gaussian noise with mean value 0 and variance 0.1 is used. The first term in Equation (1) represents regulation function f(z):

$$f(z) = \frac{1}{1+e^{-z}} ; z = \sum_{u_{sd}} W(u_{ij}, u_{sd}) \cdot x_{u_{sd}} - \theta_{u_{ij}} \quad (2)$$

Dynamically adjusting $\theta_{u_{ij}}$, the number of lightpaths, assigned to node pair can be controlled. $\theta_{u_{ij}}$ is defined as follows (Koizumi et al., 2008):

$$\theta_{u_{ij}} = -\frac{y_{u_{ij}} - y_{min}}{y_{max} - y_{min}} \cdot 2\theta_c - \theta_c \quad (3)$$

where $y_{u_{ij}}$ – load on link between nodes i and j,
y_{max} - maximum load on link in the network,
y_{min} - minimum load on link in the network,
θ_c – coefficient, which scales y_{max} and y_{min}.

The amount of lightpaths, assigned to node pair u_{ij} is a function of unit control value $x_{u_{ij}}$. More lightpaths are assigned to node pair, which has higher control value.

Regulatory matrix $W(u_{ij}, u_{sd})$ is the most important parameter of Equation (1), since it shows the relationships between node pairs. Every element of this matrix can be -1, 0, or 1; it represents the influence of node pair u_{ij} on node pair u_{sd}. According to model, proposed in (Furusawa et al., 2008), -1 corresponds to inhibition of the node u_{ij} by u_{sd}, 0 corresponds to no relation and 1 to activation. These values are multiplied by corresponding control values u_{ij}. In such way node pairs affect each other – if u_{sd} is activated by u_{ij}, increasing in $x_{u_{ij}}$ will increase $x_{u_{sd}}$ and, as it was mentioned above, increase the amount of lightpaths between nodes *s* and *d*. In (Furusawa et al., 2008) has been shown that if node pairs can activate or inhibit each other with probability 0.03 and no relation with probability 0.94, such network will be extremely adaptable to changes in traffic demand. For example, if your network consists of 6 node pairs, one of possible regulatory matrix realizations can be found in Table 1. As it can be seen, node pair u_{12} is inhibited by node pairs u_{23} and u_{24}; u_{14} is activated by node pair u_{23}.

Table 1

Example of the regulatory matrix

Node pair	u_{12}	u_{13}	u_{14}	u_{23}	u_{24}	u_{34}
u_{12}	0	0	0	0	0	0
u_{13}	0	0	0	0	0	0
u_{14}	0	0	0	0	0	0
u_{23}	-1	0	1	0	0	0
u_{24}	-1	0	0	0	0	0
u_{34}	0	0	0	0	0	0

So for every node pair we construct regulation function, which is defined by Equation (2). Adjusting $\theta_{u_{ij}}$, we can move regulation function in the negative or positive direction, as it is shown in Fig. 1.

As it was mentioned above, v_g indicates conditions of VNT. As an input parameter for v_g maximum link utilization in the network is used. Maximum link utilization is selected as an input parameter because it can be easily retrieved using Simple Network Management Protocol (SNMP). But other parameters, such as average delay between node pair u_{ij} can also be used to indicate the conditions of VNT. It will be referred to v_g as *activity* or *goodness* of VNT after this. Activity of VNT is defined as follows (Koizumi et al., 2008):

$$v_g = \begin{cases} \frac{100}{1+e^{(\delta\cdot(u_{max}-\zeta))}}, & u_{max} \geq \zeta \\ \frac{100}{1+e^{(\frac{\delta}{2}\cdot(u_{max}-\zeta))}}, & u_{max} < \zeta \end{cases} \quad (4)$$

where δ - gradient of link utilization,
ζ – target link utilization.

From Equation (4) it can be seen that ζ is a threshold for VNT activity v_g. If maximum link utilization is more than ζ, activity of VNT dramatically degraded and stochastic behavior dominates over deterministic and the system is searching for a new attractor. Activity function can be seen in Fig. 2.

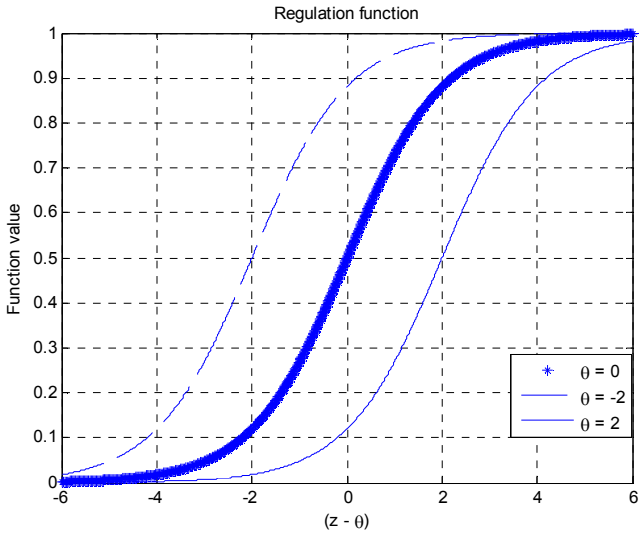


Fig. 1. **The regulation function's example.**

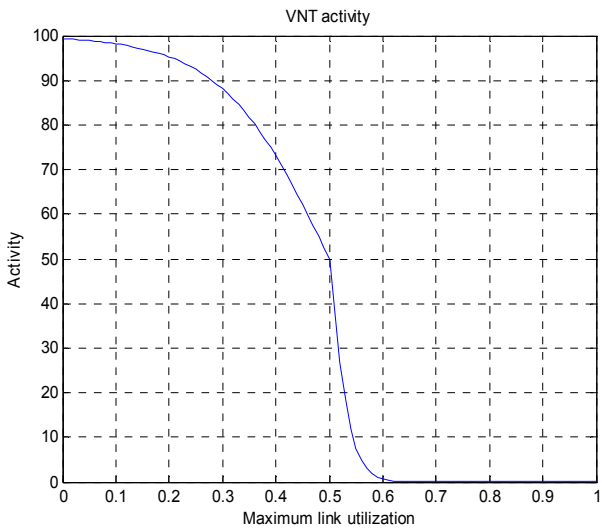


Fig. 2. **VNT activity function.**

As it was mentioned above, the number of lightpaths between node pairs depends on control unit value $x_{u_{ij}}$. More lightpaths are assigned to node pairs with a higher control unit value. It is supposed that every node has K_r receivers and K_t transmitters.

The amount of assigned transmitters and receivers is a function of $x_{u_{ij}}$ normalized by all control unit values of all node pairs, which use transmitters and receivers from node *i* to node *j*. The number of lightpaths between nodes *i* and *j* is defined as follows (Koizumi et al., 2008):

$$N_{u_{ij}} = \min\left(K_r \cdot \frac{x_{u_{ij}}}{\sum_s x_{u_{sj}}}\right), \min\left(K_t \cdot \frac{x_{u_{ij}}}{\sum_d x_{u_{id}}}\right) \quad (5)$$

Network simulations

The network with 6 nodes and 15 bidirectional links will be simulated. Each node has 6 transmitters and 6 receivers. Traffic demand matrices are randomly generated in certain time moments *T*=10, 20, 40, 60, 90. The

total time of simulation is 110. In these simulations two parameters – ζ and δ will be changed and it will be analyzed how it affect VNT activity and maximum link utilization.
In the first set of simulations δ is set to 50 and ζ to 0.5. The results of simulation – maximum link utilization and VNT activity – can be seen in Fig. 3 and Fig.4 respectively.

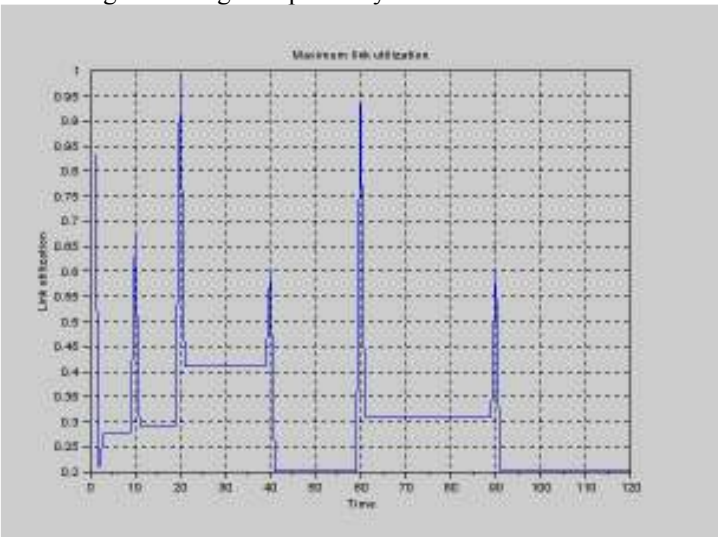


Fig. 3. **Maximum link utilization.**

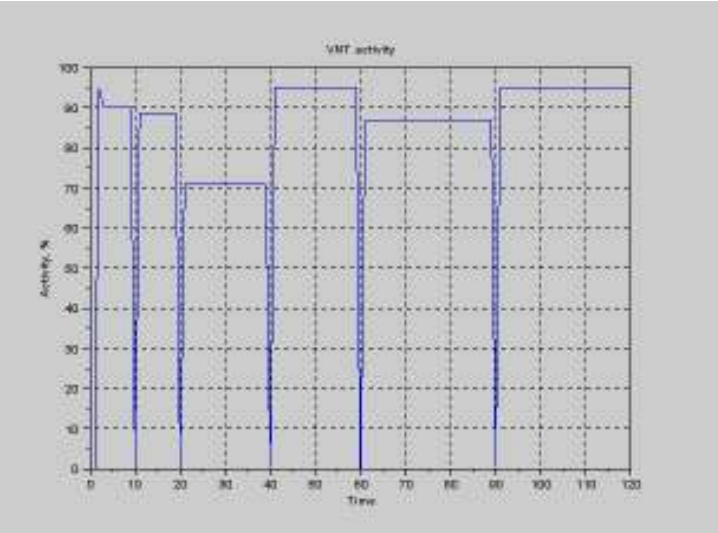


Fig. 4. **VNT activity.**

In time slots T=10, 20, 40, 60 and 90 VNT activity degrades due to changes in environment (maximum link utilization increases), in result stochastic behavior dominates system and the system state is driven by η. System randomly fluctuates searching for a new attractor. As soon as the system is close to an attractor VNT activity increases and deterministic behavior again dominates over stochastic one.
As it is seen, results are good, but in time period 20-40 VNT activity is only about 70%. Let us try to improve network activity by adjusting δ. By default δ value is 50. Let us see what happens if δ will be set to 30 and 80. The results can be seen in Fig. 5 and Fig.6. By increasing δ, we achieve quicker and more precise responses to changes in maximum link utilization.
In Fig. 5 and Fig.6 we can see a very interesting effect. In time period 0 – 10 and δ=30, the system conditions are good, but the maximum link utilization increases unexpectedly (Fig. 5), although there are no changes in traffic demand at this time moment. It happens because stochastic behavior always influences system conditions, even if VNT conditions are good. It causes VNT activity degradation. At this moment noise drives the system and a new attractor is being searched. After some time a new attractor has been found and system activity recovers. Because of noise random nature it cannot be predicted when it will happen.

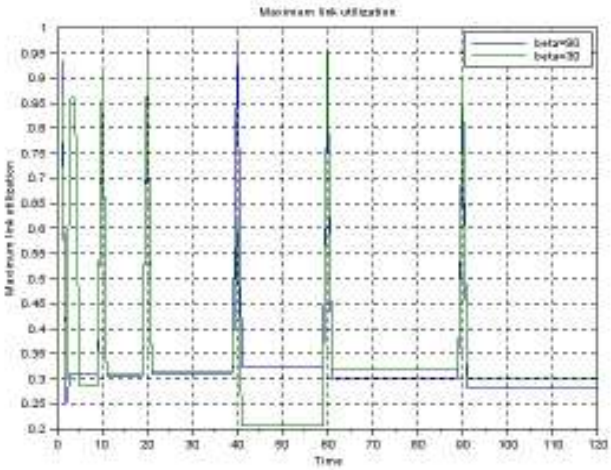


Fig. 5. **Maximum link utilization with respect to δ.**

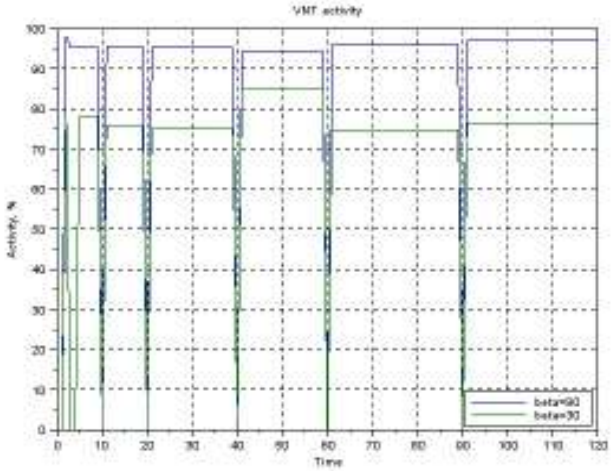


Fig. 6. **VNT activity with respect to δ.**

Now the target maximum link utilization ζ, which by default is 0.5, is changed. It is changed to 0.3 and 0.8. Results are presented in Fig. 7 and Fig. 8.

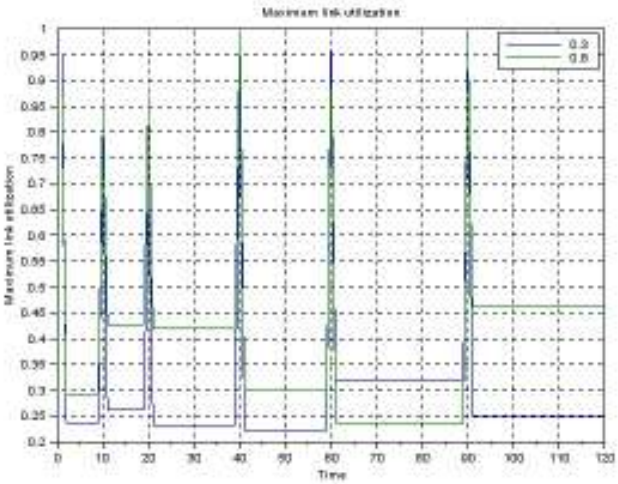


Fig. 7. **Maximum link utilization with respect to ζ.**

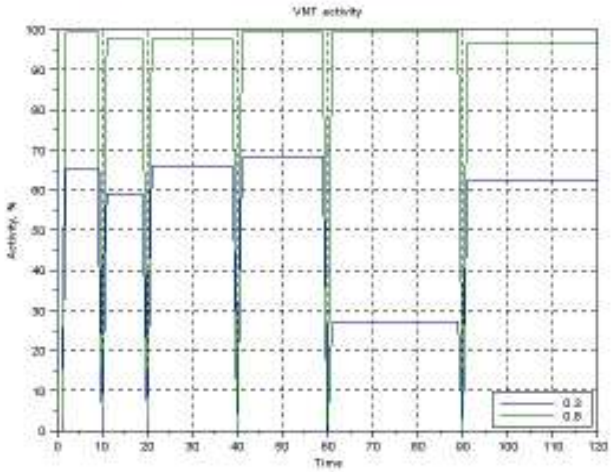


Fig. 8. **VNT activity with respect to ζ.**

In time slots T= 10, 20, 40, 60 and 90 due to changes in traffic demand and as a result changes in maximum link utilization VNT activity degrades. In these moments stochastic behavior dominates over deterministic as the first two terms in Equation (1) are zero or close to zero. Control unit values randomly fluctuate searching for a new attractor. As soon as a new attractor is found, VNT activity recovers and deterministic behavior again drives VNT.

As it can be seen from Fig. 7 and Fig. 8, if ζ is improperly selected, VNT goodness is poor even if maximum link utilization is not so high – please compare, in time interval $T\in$ [90-120] maximum link utilization and VNT activity for threshold ζ. For ζ = 0.3 maximum link utilization is only about 0.25, but VNT goodness is 63%, but for ζ = 0.8 maximum link utilization is 0.47, but VNT goodness is close to 100%. It is obvious that in this case ζ is not appropriately chosen and this affects VNT activity and in the end VNT performance.

Conclusion

- Gaussian noise cause system randomly fluctuates searching for a new attractor. At this moment stochastic behavior dominates deterministic.
- Adjusting activity gradient δ can achieve quick responses to changes in VNT conditions.
- If activity threshold ζ is improperly selected, this can cause improper VNT conditions evaluation.
- As it has been shown in some cases Gaussian noise can cause system unexpectedly change its state. But on the other hand thanks to the same Gaussian noise the system quickly re-establishes its conditions.

References

Agrawal, G. and Medhi, D., November 2006. Lightpath topology configuration for wavelength-routed IP/MPLS networks for time-dependent traffic. In Proceedings of GLOBECOM, pp. 1-5.

Chen, B., Rouskas G. N. and Dutta R., October 2008. On hierarchical traffic grooming in WDM networks. IEEE/ACM Transactions on Networking, vol. 16, pp. 1226-1238.

Furusawa, C. and Kaneko, K., January 2008. A generic mechanism for adaptive growth rate regulation. PLoS Computational Biology, vol. 4, p. e3.

Koizumi, Y., Miyamura, T., Arakawa, S., Oki, E., Shiomoto, K., Murata, M., 2008. Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Systems, Hyogo, Japan, Article No. 9.

Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E.D. and Taft, N, June 2004. Structural analysis of network traffic flows. In Proceedings of ACM Sigmetrics, pp. 61-72.

Ricciato, F., Salsano S., Belmonte A. and Listanti M., June 2002. Off-line configuration of a MPLS over WDM network under time-varying offered traffic. In Proceedings of IEEE INFOCOM, vol. 1, pp. 57-65.

A methodology for automated geosimulation model generation

Kaspars Cabs, Arnis Lektauers, Yuri Merkuryev

Riga Technical University, Meza street 1/3, Riga, LV-1048, Latvia
E-mail: kaspars.cabs@gmail.com, arnis@itl.rtu.lv, merkur@itl.rtu.lv

Abstract: *Information is one of the most important business success factors nowadays, as it provides an opportunity to respond to emerging changes and to take effective decisions. This information has to be precise and easily accessible. The aim of this research is to create a methodology for automatic real world adequate geosimulation model preparation, based on Geographical Information Systems (GIS) models and statistical data collected. All structures of an urban environment interact, and the urban environment is a network that exists among different structures, links, flows and relationships. The spatial distribution and development direction of the territory, also business environment development is determined by a street network and transport communication network. Besides, city development plans require the newest information about transport traffic and pedestrian flows as well as information on the accessibility of local services, open territories, etc. This information is well formed in GIS models. Most of model objects also contain capacity information. As the GIS models are static, they cannot be used to determine any dynamics of the territory; however, the urban environment is nonlinear. A possible way to solve nonlinearity is to prepare geosimulation models, as they are more informative for dynamic processes than GIS models. Geosimulation models also allow making experiments and solving optimization tasks. The automatic creation of geosimulation models does not require deep knowledge of simulation techniques for domain specialists further using them. From the economic aspect, it is characterized by lower expenses, quicker access to data and larger flexibility.*

Keywords: Visual interactive modelling, GIS, Business geolocalization, geosimulation.

Introduction

The 21st century is marked by rapid development of Information Technologies (IT), which is also promoting constant changes in the society, economy, environment, as well as having its impact on the territorial development and spatial planning. In many cases, changes in spatial structure of urban or rural territories create problems of social and economic nature, and most of these have to do with balanced development in the aforementioned spheres. When territory development is planned, there must be an all-encompassing analysis of the relevant spatial structure, as well as of the historical development thereof. Any development model must ensure balance among the economic, social and environmental spheres in time and space. An urban environment is a dynamic system which is constantly changing in terms of space and time. The greatest changes in the spatial structures can usually be seen when political and economic systems change. All of the structures of an urban environment interact and the urban environment is a network that exists among different structures, links, flows and relationships. The spatial distribution and development direction of the territory, also business environment development is determined by street network and transport communication network. This structure is one of those which should be regularly monitored, as it has the largest impact on our everyday activities. The coordinated street hierarchy and integrated public transport flow can significantly improve our life quality.

The analysis of spatial data is an important backbone for many applications and decision processes in research and business. Nowadays business has to undergo continued adaptation to the new circumstances, which are growing more and more dynamic in the time dimension. One of the most important business success factors is information, which gives the opportunity to respond to the changes and to take effective decisions, and in order to do so this information has to be precise and accessible, for time is money. Also city development plans require the newest information about transport traffic and pedestrian flows as well as information on the accessibility of local services, open territories etc.

The main problems addressed are availability, regularity of updating, quality of databases and integration in the system for users and its expensiveness.

Data quality is a problem, not only referring to the usual problems with the attribute data and statistical data; an additional serious challenge is the accuracy of digital boundary and network files.

Background

Geographical Information Systems (GIS) are one of the main elements for spatial analysis. GIS impact on environmental science studies, economics, prediction of urban planning and design, infrastructure development, land and population management, planning of new directions of development, monitoring, modelling, and further research is priceless (Lee, 1990; Cekule, 2007), as are digital cartography (Berry, 1987) and databases of social and economic processes. Therefore, merging these elements into a single GIS system and deploying it on Internet through a single platform, opens a wide field of opportunities for spatial process analysis and

interpretation, as well as the process of spatial analysis and modelling of real processes and phenomena becomes faster, more accessible and understandable to decision-makers (Goodchild, 2000; Torrens and Sullivan, 2001) because the use of spatial analysis methods can create good quality models of spatial division, social, economic, demographic and other models of processes and phenomena (Benenson and Torrens, 2004; Berec, 2002; Berger, 2001).

Even though GIS models ensure high degree of details, they are still static models. In order to analyse dynamics, additional modelling methods have to be used in the urban space. For this purpose often cell automata based modelling methods are used. In this case, the interaction between the cells creates a dynamic reproduction (e.g. the historic urbanisation). However, this modelling method can be used in the urban environment only if the level of detail is comparably small, and cannot be applied for modelling tasks involving details to the level of houses and streets. Urban environment usually features non-regular construction structure. In order to solve the modelling tasks requesting a high level of details, microscopic modelling methods can be used – discreet event-based modelling or the Multi-agent simulation (MAS).

Task

This paper reviews the methodology which is based on the analyses of correlations between GIS models of the territory, the existing statistical data for objects in this territory and further creation of geosimulation model. It also includes a methodology for preparation of input data and provides a description of simulation model creation process. As a result, a simulation model adequate to statistical data is created, which is ready for further experiments or optimisation tasks.

The research uses Multi-agent system as the modelling method, which is adapted to the requirements of urban environment simulation. In the example, GAMA simulation platform and model is used, which is created in GAML simulation language.

GIS

GIS is a system for mining, storing, update, manipulation, analysis and display of space- and time-related data (all forms of geographic information forms). This information is stored in the „spatial" database consisting of spatial objects, e.g. rivers, roads, borders, buildings, etc. each having its identification number. GIS information contains details on location of the object (coordination system) and its metadata, e.g. source of the data, time of its creation, exactitude, data class, etc. Still in relation to modelling time GIS is a static system thus not allowing for accomplishment of modelling tasks. From the point of view of modelling technologies, GIS forms a basis on which modelling tasks can be accomplished.

MAS

We are living in a complex world in which all elements are connected to each other. A multi-agent system (MAS) is one of solutions aiming at explaining this complexity in the same way as it exists in real world. The MAS-based simulation is a micro simulation technique, where some or all of the simulation entities are implemented as agents. In micro simulation all objects, algorithms of objects and resources are well described. In order to achieve results of modelling experiment, n iterations are necessary. There are two directions of microscopic simulation: discrete and event-based simulation; and agent-based simulation.

Table 2

Business and Organizations • Manufacturing • Consumer markets • Supply chains • Insurance Economics • Artificial financial markets • Trade networks Infrastructure • Electric power markets • Hydrogen economy • Transportation Crowds • Human movement • Evacuation modelling	Society and Culture • Ancient civilizations • Civil disobedience Terrorism • Social determinants • Organizational networks Military • Command & control • Force-on-force Biology • Ecology • Animal group behaviour • Cell behaviour • Sub cellular molecular behaviour

The multi-agent system is a set of agents in a certain environment, where these agents are linked by certain mutual relation rules.
Agents in MAS should correspond to the following principles:
Autonomy – agents must operate without discrete intervention by humans or other agents;
Social ability – agents interact with other agents;
Reactivity – agents must react to changes in the environment in time for the action to be useful;
Pro-activeness – agents must take initiative to act in their operating environment (Wooldridge and Jennings, 1995).
The usage of MAS is many areas (Table 1 – Macal and North, 2005).

Geosimulation

Geosimulation model is based on a real existing adequate model creation and conducting experiments with it. Simulation model can be described as a black box with input X and output Y (Fig.1).

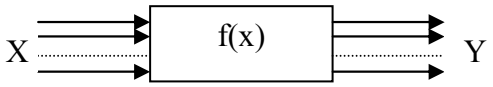


Fig. 1. **Black box.**

The creation of a geosimulation model implies finding such function f(x), which guaranties the same or close to the same result Y as in the real world. In this case the model should include such inputs X, which are significant to the result Y. In most cases the model f(x) is not linear, which means that it cannot be calculated only with mathematical methods. In experiments it is possible to simulate what happens if the input X is changed; and sometimes also change of the model f(x), e.g. to see what happens if the left turn is prohibited on a street.
Simulation environment used is a multi-agent simulation system, which allows describing each object as an agent.
According to Torrens (Benenson and Torrens, 2004), formally, a Geographic Automata System (GAS), G, may be defined as consisting of seven components:

$$G \sim (K, S, T_s, L, M, N, R_N)$$

where

K – set of types of automata featured in the GAS,
S – set of states

$T_s: (S_t, L_t, N_t) \rightarrow S_{t+1}$ – set of state transition rules, used to determine how automata states should change over time.
L – the georeferencing conventions that dictate the location of automata in the system

$M_L: (S_t, L_t, N_t) \rightarrow L_{t+1}$ – the movement rules for automata, governing changes in their location in time
N – represents the neighbours and their relations of the automata

$R_N: (S_t, L_t, N_t) \rightarrow N_{t+1}$ - rules that govern changes of automata relations to the other automata in time.
Another concept of geosimulation is when all GIS layers also are automata and GIS only represents their location. In this case simulation system adds additional parameters for each GIS layer agent, which is necessary for simulation. In this case, according to Patrick Taillandier1 (Taillandier P., Drogoul, 2010) ,integration of GIS date is agentification.

GAMA Simulation Platform and GAML language

GAMA (a GIS & Agent-Based Modelling Architecture – Fig. 2) is a multi agent simulation system. It has been developed by the research team MSI (located in the IFI, Hanoi, and part of the IRD/UPMC International Research Unit UMMISCO) since 2007. The system is strongly oriented towards the use of GIS models as simulation background. GAMA simulation platform uses the GAML simulation language. All objects in GAML are agents, where each has his own properties and activities. As an example, there are several main properties: visible – agent is visible on the screen, situated – agents located in the environment which coordinates, defined size, moving – agents are able to move in the environment, carrying – agent can contain other agents. There are also special properties which allow agents to communicate to each other, to database, exploring properties for robotic tasks. An agent can possess many properties – for example, an agent can be visible and moving. Language is structured in several parts. GLOBAL section defines all main parameters, like the quantity of agents, movement speed, etc. There are two variations of global parameters – visible (users can change parameters without changing the model) and hard defined.

.

Fig. 2. **GAMA Simulation Platform.**

The next part defines an environment where simulation takes place. Agents should be defined in "entities" section. Visual part of simulation is defined in "output" section. Both, visual reproduction of the territory itself is possible, as well as display of different charts explaining the simulation.

Agent definition means specification with one or several main properties. Agents can be object-oriented, which means that they can inherit properties from parents; and in this case child agents use parent agent properties and actions plus newly specified properties and actions.

Agents can communicate with other agents in the system by using directive "ask". If an agent is created from GIS model, each GIS model element is described in GAML as an agent where all GIS element properties stay as agent properties in GAML.

Proposed method

Typical structure of simulation project includes:

1. Problem definition. Information extraction from the environment, analysis of elements and their relationships, creation of simulation task.
2. Creation of conceptual model and simulation model in one of simulation languages.
3. Simulation model adequacy tests.
4. Experiments and analysis of simulation results.

Geosimulation model consists of static elements – houses, industrial buildings, roads, pathways, other structures, and moving objects (cars, people, etc.). Some of moving objects can carry other objects (e.g. people travelling in cars, busses). These objects move between the static objects, they may stay in other objects, leave them or move inside static objects. Simulation tasks include an analysis of flow and load.

The goal of automated geosimulation model is to speed up conceptual simulation model creation; in addition, it also lowers the necessary costs. The prerequisites for automated geosimulation model creation are:

1. Territory of geosimulation model is real; and its GIS model exists;
2. Simulation task is flow or load analysis;
3. There is free access to statistical data in any place required for the model.

The obtained simulation model is adequate to statistical data and ready for future experiments.

The proposed method consists of input data preparation, which should be done by user, and geosimulation model creation which is done automatically.

Input data preparation

The statistics database should contain information on the average movement times between static objects. Only routings defined in statistics database are allowed, other routings are not allowed in the model. In this way the allowed routes are specified, and it saves time necessary for the verification of all allowed routes in the analysed territory.

GIS models should include all static objects shown in statistics database, additional important objects and other visual objects necessary for better representation of simulation model.

Rules specify allowed access times for static carrying object. Rules format is Extensible Markup Language (XML). XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

Geosimulation model creation

Model creation is based on patterns. Models are already prepared as GAML patterns, in which the changeable places have to be replaced by information on the GIS model objects. The same system is applied for objects defined in rules xml, GAML actions and reflex (events)– all actions and events are already described in patterns. Only when all input data are ready, a model can be created. Model creation process includes several steps:

1. Creation of initial model in GAML from pattern.
2. Modification of the global section in initial model (creation of specifications for moving objects and specifications for static objects, defined in rules xml. Creation of links to GIS models.
3. Creation of actions for the static objects, specified in rules xml. These actions are created using action patterns, and the creation process is similar to the model creation process. First, the initial action and reflex parts for object and necessary variables parts are created, as they should be included in GAML at the beginning of entities section. Then replacement with the original coding from rules xml takes place (in order to name an object, 'object id has to be added to the object name in GAML).
4. Exporting the obtained model to the user.

Example

An example of the prototype of RTU Campus is presented below, for which the GAML simulation model had to be created. In this example the first task was creation of a GIS model from the territory visualisation model (see Fig. 3). For this purpose a simplified GIS model was created showing the existing buildings, streets between the buildings and borders of the territory. Each building is indexed (building1, building2 … , building17). In the sample also specific additional parameters are added for the buildings, e.g. building type – residential building or a building for studying, building containing cafeteria, and its capacity. Streets in the sample have unlimited capacity. As this is a theoretical sample, the next step for this case is creation of a statistical database. The statistical database should show paths from one object to another, the average times and percentage of the the route usage (Table 2 – Statistical database sample).

Fig. 3. **RTU Territory visualisation and GIS representation.**

Table 2

Statistical database sample

ID	Object from	Object to	Average time	Percentage
1	Building15	Building11	3:03	12%
2	Building15	Building5	5:24	4%
3	Building15	Building7	5:11	2,3%

The next step is rules XML creation. It should contain static object, allowed/denied times and usage percentage of allowed times. In rules XML ID tag with object code and allow/denied tags should exist:

```
<rules>
      <rule>
          <id>Building15</id>
          <allowfrom>00:00</allowfrom>
          <allowto percentage="50">08:00<allowto>
          <allowto">09:10<allowto>
          <allowfrom percentage="30">14:30</allowfrom>
          <allowfrom >16:30</allowfrom>
          <allowto>24:00</allowto>
    </rule>
</rules>
```

Tags <allowfrom> and <allowto> contain additional index "percentage". After the tag of this form there always should be another same tag with or without "percentage" index, but the last tag always should be without index. This means that it is allowed for a certain percentage of object capacity to come or to leave this building. In GAML code classical random function is used to choose which moving object applies to this criterion. In the sample for street objects there are no rules, so they are used only as paths along which the movement takes place with Djikstra optimisation:

```
set the_graph <- (as_edge_graph(road as list)) with_optimizer_type "Dijkstra";
```

When rules XML is ready, model generator is ready to prepare GAML code. After GAML code generation it is possible to set up in GAMA environment the moveable objects (or agents according GAMA). The result should be similar to that shown in Fig. 4

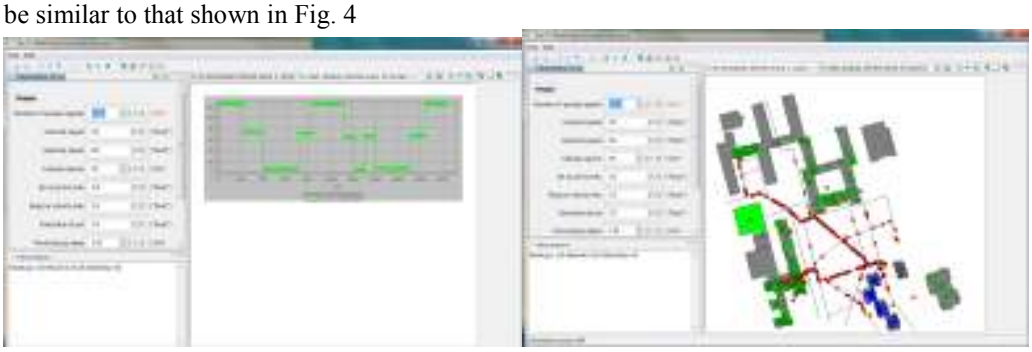


Fig. 4. **GAMA Example.**

With the generated code, users can made further experiments, change model rules by changing the GAML code. The model is adequate to the statistical data, so each movement time from one object to other is close to movement time in the statistical database. For speed calculation there is used Normal distribution to speed-up the calculation.

Conclusions

The presented research is devoted to the exploitation of the multi agent systems and their correlation with GIS models, and this type of models are geosimulation models. One of geosimulation solutions is “agentization” of GIS models, where all GIS model objects are interpreted as agents in geosimulation model.

The creation of geosimulation model is time-consuming and complicated. It is possible to automate many functions of model creation. Using statistical data and capacity data in GIS models, and additionally describing movement object nature in static objects (houses, roads) allows for the automated model creation. The obtained models are adequate to statistical data and described in the rules document, and are ready for future experiments.

The described system is part of a complex solution, which includes also gathering of statistical information and its preparation for future tasks in model creation. Collection of statistics is automated by statbox units, which take video of analysed territory, this information is then analysed in the statbox server, recognized in other statbox units, and based on the recognized information the statistical database is created. The use of this complex system allows one to solve many geolocalization tasks.

References

Baumanis, I., Gailis, A., Liepiņš, K., 2002. Priežu sēklu plantāciju pēcnācēju novērtējums (Evaluation of progeny of Scots pine seed orchards). *Mežzinātne*, 45, pp.46-59. (In Latvian)

Benenson, I. and Torrens, P.M.,2004. Geosimulation: Automata-based Modeling of Urban Phenomena. 2004 John Wiley & Sons, Ltd ISBN: 0-470-84349-7

Berec, L., 2002. Technigues of spatially explicit individual-based models: construction, simulation, and mean-field analysis. Ecological Modelling. 150. pp.55-81.

Berger, T., 2001. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes, and policy analysis. Agricultural Economics. 25. pp.245-269.

Berry, J.K., 1987. Fundamental operations in computer assisted map analysis. Journal of Geographic Information Systems. Vol.1, No.2. pp.119-136.

Braksh, H.A., 1971. Sapropelevye otlozhenye i puty ih ispolzovanya (Sapropel sediment and methods of its application). *Riga –Zinatne*, pp. 248-258.

Cabs, K. and Cekule, M., 2012. StatBox simulation system concept for spatial analysis, 15th IFF Science Days in Magdeburg, 9th Digital Engineering and Virtual Technologies Conference, Magrebourg, Germany, 2012.

Cabs, K., Lektauers, A., Merkuryev, Y., 2011. Visual Interactive Modelling of Business Geolocalization Annual International Conference on "Virtual and Augmented Reality in Education" Vidzeme University of Applied Sciences, Valmiera, Latvia, 2011.

Cekule, M., Zīra, M., Čabs, K., 2007. Interactive business geolocalization. XXIII International Cartographic Conference Proceeding. Moscow

Goodchild, M.F., 2000. The Current Status of GIS and Spatial Analysis. Journal of Geographical Information Systems. Vol.2, pp 5-10.

Macal, C.M. and North, M.J., 2005. TUTORIAL ON AGENT-BASED MODELING AND SIMULATION. Proceedings of the 2005 Winter Simulation Conference, M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

Lee, Y.C., 1990. Geographic information system for urban applications: problems and solutions. Environment and Planning B: Planning and Design 17, 463-473.

Taillandier, P., Drogoul, A., 2010. From GIS Data to GIS Agents, Modeling with the GAMA simulation platform. 2010, http://www.irit.fr/TFGSIM/DOCS/PatrickTaillandier_GAMASimulationPlatform.pdf Technical Forum Group on Agent and Multi-agent-based Simulation, 1st meeting co-located with Eumas'10 in Paris, France

Torrens, P.M., O'Sullivan, D., 2001. Cellular automata and urban simulation:where do we go from here? Environment and Planning B. 28(2), pp. 163-168.

Wooldridge, M and Jennings, N. R., 1995. Intelligent agents: theory and practice. The Knowledge Engineering Review / Volume 10 / Issue 02 / June 1995, pp. 115-152.

Application of databases for development of stoichiometric and dynamic models of biochemical networks

Natalja Bulipopa, Ilona Odzina

Latvia University of Agriculture, Liela Str. 2, Jelgava, LV-3001, Latvia
cvetkova.natalja@gmail.com

Abstract: *Metabolic engineering uses tools of bioinformatics, systems biology and synthetic biology to modify organisms for new biotechnological tasks. Stoichiometric and dynamic models are needed for successful design of requested bioprocess. In spite of available model databases both for stoichometric and dynamic models they often are applicable just partly as deposited models may not cover the organism and process of interest. Depending on that intensive use of databases about particular reactions and metabolites may be required. This study analyses and suggests possible application of different publically available databases developing stochiometric and dynamic models. The contents and applications of data bases KEGG, BRENDA, SABIO-RK, IntEnz, MetaCyc, ChEBI, PubChem, BioModelsare analyzed.*

Keywords: metabolic engineering, KEGG, BRENDA, SABIO-RK, BioModels.

Introduction

Metabolic engineering is a growing field of research in microbiology serving as an integrated approach to design new cell factories by providing rational design procedures and valuable mathematical and experimental tools (Patil et al., 2004). Different aspects of bioinformatics, systems biology and synthetic biology are exploited to improve the features of cell factories (Nielsen and Keasling, 2011). In the field of bioinformatics, systems biology and synthetic biology one of the main research challenges are the understanding of biological and biochemical process, integration of existing data in databases this contributing to the analysis of the use of new analytical methods. The database provides integrated tools, enabling the user to independently perform data analysis.

Both stoichiometric and dynamic models are used in metabolic engineering. The development of stoichiometric models or stoichiometric reconstructions is generally well developed (Thiele and Palsson, 2010) and there are several data bases offering published models. Still important organism specific information has to be implemented in the model by biologists specializing in the particular organism (Pentjuss et al., 2013) because comparison of different models of the same organism often demonstrates high disagreement (Mednis and Aurich, 2012). In case of dynamic modeling there are dynamic models available in dedicated database (Novère et al., 2006). Still dynamic models are relatively small in terms of number of involved reactions (usually some tens of reactions) and often they are not covering the process of interest or it is modeled under different circumstances. Therefore data bases are needed to find very specific kinetic parameters to perform parameter estimation and optimization tasks (Mendes and Kell, 1998; Sulins and Mednis, 2012; Kostromins et al., 2012).

Biological databases are an important tool in assisting scientists to understand biological phenomena from the structure of biomolecules and their interaction, to the whole metabolism of organisms and to understanding the evolution of species. This knowledge facilitates the fight against diseases, assists in the development of medications and in discovering basic relationships amongst species in the history of life (Makheswari and Sudarsanam, 2012).

In case of metabolic engineering the first step is to set the task of the cell factory: substrate, product and limitations of the task (Lee et al., 2010). After that the choice of the organism of interest has to be done. The next challenge is to choose for each analysis stage corresponding initial acquisition of information or database.

The aim of this article is to analyze the workflow of computer modeling in metabolic engineering of biological organism and database usage.

Computer modeling in metabolic engineering of biological organism

Before choosing a database, at the beginning must be defined an objective (task) – to set the task of the cell factory: substrate, product and limitations of the task.

When the objective (task) was defined, must choose the organism and some process of chosen organism which will be analyzed and the structural model of chosen organism will be developed. The structural model contains reactions, their directions and metabolites. For acquisition of this initial information, the following free of charge database are often used: Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2006), BRENDA (Chang et al., 2009) and SABIO-RK (Rojas et al., 2007).

The created structural model allows us to analyze and illustrate the network of biochemical processes – cycles and node degrees (Fig. 1) (Rubina, 2012; Odzina et al., 2010).

This model describes interaction of the reaction and metabolites. In order to create organism's reconstruction the additional information is needed about the reaction's genes, proteins, subsystems, enzyme numbers, metabolites and their full title, neutral formula, charged formula, charge, compartment, ID numbers and encoding (InChi string and SMILES). To obtain information simultaneously can be used several databases: KEGG, BRENDA, IntEnz (Integrated relational Enzyme database) (Alcántara et al., 2013), MetaCyc (Caspi et al., 2012), BioCyc (Caspi et al., 2010), Chemical Entities of Biological Interest (ChEBI) (Matos et al., 2010), and PubChem (Bolton et al., 2008). All this information is necessary for the analysis and inspection of reconstruction by COBRA Toolbox software (Schellenberger et al., 2011).
If after the reconstruction of model development has been ascertained that in the model is interaction between all reactions and metabolites, then can analyze the dynamic parameters of the model. Thus, the dynamic model is created. Model creation is necessary to add information of the kinetic equations and parameters. The following databases: BRENDA (Chang et al., 2009), SABIO-RK (Rojas et al., 2007), and BioModels (Novère et al., 2006) contain most of necessary information.
The last step after the dynamic model development is parameter estimation, based on a dynamic model simulation data. When, model's simulated data are approximated to experimental data, then developed dynamic mode (*In silico*) can be used in laboratory experiments (*In vitro*), in order to reduce the number of failed experiments.

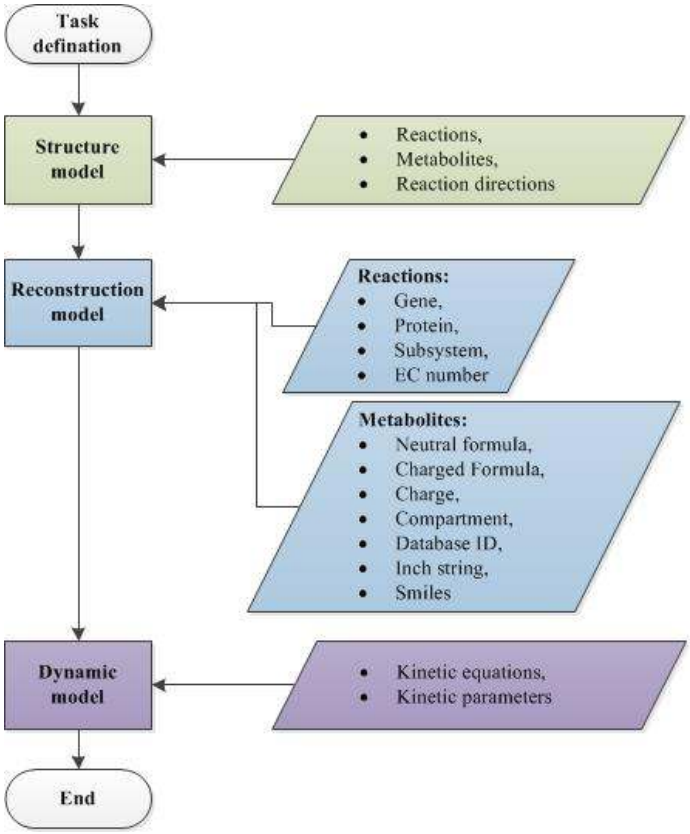


Fig. 1. **Algorithm of metabolic engineering computer modeling of biological organism.**

Biological Databases

Kyoto Encyclopedia of Genes and Genomes (KEGG): is a bioinformatics database containing information on genes, proteins, reactions and pathways. The organisms section is divided into eukaryotes and prokaryotes, encompasses many organisms for which gene and DNA information can be searched by typing in the enzyme of choice. This resource can be extremely useful when building the association between metabolism enzymes, reactions and genes (Ogata et al., 1999). As well as for overall examination of enzymatic reactions and EC numbers the database KEGG can be used. KEGG EC number information and reaction directionality are not organism – specific (Kanehisa et al., 2006).
BRENDA: a comprehensive enzyme database. Database contains functional data for all enzyme classes (~ 4800 entries in six main classes in 2008) that have been classified according to the EC scheme of the IUBMB (International Union of Biochemistry and Molecular Biology) irrespectively of the enzyme's source. The range of data in database is not restricted to specific aspects but includes a wide area of biochemical and molecular properties of enzymes such as a) classification and nomenclature; b) reaction and specificity; c) functional parameters; d) organism-related information; e) enzyme structure; f) isolation and preparation; g) literature

references; h) application and engineering; i) enzyme–disease relationships. All data and information are manually extracted from the primary literature and are connected to the biological source of the enzyme, i.e. the organism, the tissue, the subcellular localization and/or the protein sequence (if available) (Chang et al., 2009). Database also, allows searching for an enzyme by name or EC number.

Organism – specific information was collected mainly from previously published data and the BRENDA database, where each reaction EC number and reaction directionality was checked and validated.

SABIO-RK - SABIO-RK is a web-accessible curated database offering information about biochemical reactions and their kinetic properties. It integrates information about reactions, such as reactants, effectors, and catalyzing enzymes, with information about organisms, tissues and cellular locations where the reactions take place, and with the kinetic properties of these reactions (type of the kinetic mechanism, modes of inhibition or activation and rate equations together with their parameters and measured values). As kinetic constants highly depend on environmental conditions used for their determination these are given together with the description of the kinetics. This also facilitates the comparison of data sets based on experiments assayed under similar experimental conditions. The database is populated by merging data from several sources. The general information about the reactions is mainly obtained from external databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes). In contrast, the kinetic data along with descriptions of the experimental conditions under which they were determined are primarily manually extracted from literature and curated by a team of scientists (Rojas et al., 2007).

IntEnz (Integrated relational Enzyme database) (Alcántara et al., 2013) is a freely available resource focused on enzyme nomenclature. A relational database provides better interoperability with other bioinformatics resources. All the major biological databases at the EBI, such as EMBL Nucleotide Sequence Database, UniProt and MSD, are relational. Enzyme portal (http://www.ebi.ac.uk/enzymeportal) to provide wealth of information on enzymes from multiple in-house resources addressing particular data classes: protein sequence and structure, reactions, pathways and small molecules. The fact that these data reside in separate databases makes information discovery cumbersome (Alcántara et al., 2013).

MetaCyc is a database of nonredundant, experimentally elucidated metabolic pathways (Caspi et al., 2012). MetaCyc contains more than 1928 pathways from more than 2263 different organisms, and is curated from the scientific experimental literature. MetaCyc contains pathways involved in both primary and secondary metabolism, as well as associated compounds, enzymes, and genes. MetaCyc data can be accessed in several ways: a) search for pathways, enzymes, reactions, and metabolites through this Web site; b) install MetaCyc on your computer in conjunction with the Pathway Tools software for faster access and more query options. This local installation also allows you to query MetaCyc programmatically using Java or PERL programs; c) Download MetaCyc data files (Caspi et al., 2012).

BioCyc is a collection of 2038 Pathway/Genome Databases (PGDBs) (Caspi et al., 2010). Each PGDB in the BioCyc collection describes the genome and metabolic pathways of a single organism. The BioCyc Web site contains many tools for navigating, visualizing, and analyzing these databases, and for analyzing omics data, including the following: a) genome browser; b) display of individual metabolic pathways, and of full metabolic maps; c) visual analysis of user-supplied omics datasets by painting onto metabolic maps, regulatory maps, and genome maps; d) store groups of genes and pathways in your account; share, analyze, transform those groups; e) comparative analysis tools (Caspi et al., 2010).

Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds (Matos et al., 2010). The term 'molecular entity' refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity. The molecular entities in question are either product of nature or synthetic products used to intervene in the processes of living organisms.

In order to create ChEBI, data from a number of sources were incorporated and subjected to merging procedures to eliminate redundancy.

Four of the main sources from which the data are drawn are: a) IntEnz – the Integrated relational Enzyme database of the EBI. IntEnz is the master copy of the Enzyme Nomenclature, the recommendations of the NC-IUBMB on the Nomenclature and Classification of Enzyme-Catalysed Reactions; b) KEGG COMPOUND – One part of the Kyoto Encyclopedia of Genes and Genomes LIGAND database, COMPOUND is a collection of biochemical compound structures; c) PDBeChem – The service providing web access to the Chemical Component Dictionary of the wwPDB as this is loaded into the PDBe database at the EBI; d) ChEMBL – A database of approximately 500,000 bioactive compounds, their quantitative properties and bioactivities, abstracted from the primary scientific literature. It is part of the ChEMBL resources at the EBI (Matos et al., 2010).

PubChem is an open repository for experimental data identifying the biological activities of small molecules. The primary aim of PubChem is to provide a publicē on - line resource of comprehensive information on the biological activities of small molecules accessible to molecular biologists as well as computation al and medicinal chemists. PubChem contents include more than: 1,000 bioassays, 28 million bioassay test outcomes, 40 million substance contributed descriptions, and 19 million unique compound structures contributed from over

70 depositing organizations. PubChem provides a significant, publicly accessible platform for mining the biological information of small molecules (Bolton et al., 2008).

BioModels database part of the international initiative BioModels.net, provides access to published, peer-reviewed, quantitative models of biochemical and cellular systems. Database aims are as follows: a) to define agreed-upon standards for model curation, b) to define agreed-upon vocabularies for annotating models with connections to biological data resources and c) to provide a free, centralized, publicly accessible database of annotated, computational models in SBML and other structured formats. Database is an annotated resource of quantitative models of biomedical interest. Models are carefully curated to verify their correspondence to their source articles. They are also extensively annotated, with a) terms from controlled vocabularies, such as disease codes and Gene Ontology terms and b) links to other data resources, such as sequence or pathway databases.

The models can currently be retrieved in the SBML format, and import/export facilities are being developed to extend the spectrum of formats supported by the resource (Novère et al., 2006).

Conclusion

There are many and varied biological databases and its containing different information from very specific to more comprehensive. But to be able to choose database, first of all, must be defined the initial task. For each model development step it is necessary to choose the most appropriate database to achieve better results. However the same databases can be used several times during the study, such as, KEGG or BRENDA, because the information in the database are ranked by biological factors, such as Metabolic Pathway Databases, Protein Databases, genomic databases et al.. Search in several databases at the same time can increase the probability that creating a computer model can be made a mistake. Consequently, the database information is appropriate to take instead of a single biological experiment, but from several databases.

When model's simulated data are approximated to experimental data, then developed dynamic model (*In silico*) can be used in laboratory experiments (*In vitro*), in order to reduce the number of failed experiments.

Reference

Alcántara, R., Onwubiko, J., Cao, H., Matos, P., Cham, J.A., Jacobsen, J., Holliday, G.L., Fischer, J.D., Rahman, S.A., Jassal, B., Goujon, M., Rowland, F., Velankar, S., López, R., Overington, J.P., Kleywegt, G.J., Hermjakob, H., O'Donova,n C., Martín, M.J., Thornton, J.M. and Steinbeck C., 2013. The EBI enzyme portal, *Nucleic Acids Res, 41*, Database issue D773–D780. doi: 10.1093/nar/gks1112

Bolton, E.E., Wang, Y., Thiessen, P.A. and Bryant, S.H., 2008. PubChem: Integrated Platform of Small Molecules and Biological Activities. Chapter 12. In *Annual Reports in Computational Chemistry*, *4*, American Chemical Society, Washington, DC.

Caspi, R., Altman, T., Dale, J.M., Dreher, K, Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Paley, S., Popescu, L., Pujar, A., Shearer, A.G., Zhang, P. and Karp, P.D., 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 33* (19), Database issue 6083-6089. doi: 10.1093/nar/gki892

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, O., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P. and Karp, P.D., 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 40 ,* Database issue *D742-D753.* doi: 10.1093/nar/gkr1014

Chang, A., Scheer, M., Grote, A., Schomburg, I. and Schomburg, D., 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, *37*, Database issue D588–D592, doi:10.1093/nar/gkn820

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M., 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, *34*, Database issue D354–D357, doi:10.1093/nar/gkj102/.

Kostromins, A., Mozga, I. and Stalidzans, E., 2012. ConvAn: a convergence analyzing tool for optimization of biochemical networks. *Biosystems*, *108*(1-3), pp. 73–77. doi:10.1016/j.biosystems.2011.12.004

Lee, F.C., Pandu Rangaiah, G. and Lee, D.Y., 2010. Modeling and optimization of a multi-product biosynthesis factory for multiple objectives. *Metabolic Engineering*, *12*(3), pp. 251–267. doi:10.1016/j.ymben.2009.12.003

Makheswari, U.M. and Sudarsanam, D., 2012. A Review on Bio Informatics for Diabetic Mellitus. *International Journal of Pharma Sciences and Research (IJPSR), **3*** (6), ISSN : 0975-9492, pp 389 - 395

Matos, P., Alcantara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S. and Steinbeck, C., 2010. Chemical Entities of Biological Interest: an update. *Nucleic Acids Research, 38*, Database issue D249–D254, doi:10.1093/nar/gkp886

Mednis, M. and Aurich, M. K., 2012. Application of string similarity ratio and edit distance in automatic metabolite reconciliation comparing reconstructions and models. *Biosystems and Information Technology*, *1*(1), pp. 14–18. doi:10.11592/bit.121102

Mendes, P. and Kell, D., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, *14*(10), pp. 869–883. doi:10.1093/bioinformatics/14.10.869

Nielsen, J. and Keasling, J.D., 2011. Synergies between synthetic biology and metabolic engineering. *Nature Biotechnology*, *29*(8), pp. 693–695. doi:10.1038/nbt.1937

Novère, L.N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L. and Hucka M., 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, *34*, Database issue D689–D691, doi:10.1093/nar/gkj092

Odzina, I., Rubina, T., Rutkis, R., Kalnenieks, U. and Stalidzans, E., 2010. Structural model of biochemical network of Zymomonas mobilis adaptation for glycerol conversion into bioethanol. In Proceedings of Applied Information and Communication Technologies 2010, pp. 50-54. ISBN 978-9984-48-022-0

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 27(1), pp.29-34.

Patil, K.R., Akesson, M. and Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. *Current Opinion in Biotechnology*, *15*(1), pp. 64–69. doi:10.1016/j.copbio.2003.11.003

Pentjuss, A., Odzina, I., Kostromins, A., Fell, D., Stalidzans, E. and Kalnenieks, U., 2013. Biotechnological potential of respiring *Zymomonas mobilis*: a stoichiometric analysis of its central metabolism. *Journal of Biotechnology*, *165*(1), 1-10 . doi:10.1016/j.jbiotec.2013.02.014

Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A. and Wittig, U., 2007. SABIO-RK: a database for biochemical reactions and their kinetics. *BMC Systems Biology*, ***1*** (Suppl 1):S6, doi: 10.1186/1752-0509-1-S1-S6

Rubina, T., 2012. Tools for analysis of biochemical network topology. *Biosystems and Information Technology*, *1*(1), pp. 25–31. doi:10.11592/bit.121101

Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A, Lewis, N.E, Rahmanian, S, Kang, J, Hyduke, D.R, Palsson, B.Ø., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), pp. 1290–1307. doi:10.1038/nprot.2011.308

Sulins, J. and Mednis, M., 2012. Automatic termination of parallel optimization runs of stochastic global optimization methods in consensus or stagnation cases. *Biosystems and Information Technology*, *1*(1).

Thiele, I. and Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), pp. 93–121. doi:10.1038/nprot.2009.203

Conceptual model and concepts for ERP system change implementation planning

Igors Mikulovs, Janis Grabis

Tele2 AB, Riga Technical University, Meza street 1/3, Riga, LV-1048, Latvia
igors.mikulovs@tele2.com, grabis@rtu.lv

Abstract: *The paper investigates the existing software change management approach at Tele2 and uses accumulated experiences as a basis for conceptualizing the planning of changes in packaged applications. Tele2 is one of Europe's leading telecom operators, offering mobile services, fixed broadband and telephony, data network services, cable TV and content services with 37,6 million customers in 11 countries. IT operations and services are being supported within Tele2.Change management, especially scheduling of changes, is known to be one of the most challenging problems in managing of IT operations. The task of finding an optimal change implementation plan is difficult considering many variables which influence the decision on resource allocation to implementation of change requests. Many IT companies do planning manually, re-allocating and re-planning resources constantly. This paper presents the results from in-depth study of change management and planning practices with the aim to identify concepts and their attributes influencing the change implementation plan. The conceptual model is elaborated based on the collected concepts. The objectives of change planning such as implementation cost and time minimization are also represented in the conceptual model. The one of the most important features represented in the conceptual model is re-planning of change implementation activities in response to changing development priorities and external events such as new updates of the packaged software. The conceptual model and concepts are validated against four similar studies. Construction of the conceptual model also allowed to identify several directions for future improvement of change management and scheduling practices. The described conceptual model will serve as a basis for further work on design and implementation of change planning decision support system.*

Keywords: Change management, conceptual model, change implementation planning, release planning.

Introduction

Enterprise resource planning (ERP) applications were one of the fastest growing and most profitable areas of the software industry during the late 1990s (Sprott , 2000). Companies worldwide have invested billions of dollars in implementing these systems (Bailor et al., 2006). The adoption of ERP systems has been the focus of substantial research in recent years (Chuck et al., 2010). Although early research has covered many aspects of ERP implementation, it has focused mainly on the implementation stages of the ERP lifecycle. As with other types of information systems, many maintenance activities must be carried out, and issues need to be resolved after an ERP system has become operational (Daveport, 1988).

Corrective, adaptive, preventive and perfective (Lientz et al., 1981) changes should be introduced during the ERP maintenance, and these changes are frequently initiated by Change Requests (CR). Implementation of the change is driven by a number of factors, for example: priority, business case, planned production dates, user needs. At some point of time there could a large number of outstanding CR, however implementation of those is always limited by resources and other constraints.

The actual problem is to plan and distribute resources (human, time, financial) over the input CR list and to elaborate a time plan and release date for particular CR list considering different factors, which impacts decision on CR processing order. Stakeholders might be interested in quickest as possible implementation, or cost saving implementation, or only interested in precious CR release date communication. Currently many companies drive planning manually, re-allocating and re-planning resources when new CR priorities are changes, resource pool is changed, new urgent CR incomes, or low priority becomes not relevant. Due to changes in resources or CR priorities, some of CR can mismatch requested production dates should be reported and confirmed to next release date. Re-allocating work plan there should be considered already performed work, and interruption penalties when specialist break work in particular activity with following return to it. All this however becomes unmaintainable at certain point of time, due to significant amount of parameters to consider, downtime of the work increases, some activities becomes uncontrolled, as result overall efficiency of planned activities drops, IT companies do not perform anymore in optimal way.

The CR list needs to fulfil the interests of various stakeholders and takes many variables into consideration. Several scholars have presented lists of such variables, including: importance or business value, stakeholder preference, cost of development, requirement quality, development risk and requirement dependencies (Carlshamre, 2002; Firesmith, 2004; Greer et al., 2004; Weerd et al., 2006).

In order to automate this task, an algorithm for allocation of the CR list to available resource subject to the relevant factors is required. That would allow change/release managers to reduce manual work, to make more

efficient decision and to better control the CR flow over the implementation period. The valuable benefit would be a possibility to communicate correct production dates for the CR despite changes of the initial plan and other variations.

This paper aims to establish a common basis for future development of this CR allocation algorithm. That comprises definition of fundamental concepts utilized in the change planning during the software maintenance. This model is a result of study of the manual change request planning within Tele2 IT division. The subject researched organization has the market-driven requirement engineering processes (Carlshamre et al., 2001). The paper also defines requirements for CR processing list and discusses factors affecting ERP system change implementation. The conceptual model can assist in CR development; add value to automated change request planning. It can provide guidance to organization on building or improving their change management process.

Methodology

The analysis process consisted of two stages. Initially, we identified common properties of the system and process studied by conducting brainstorming sessions and collecting relevant data. CR planning concepts were individually analysed, and added to a draft conceptual model. All information was gathered using different sources, change/release planning documentation, system specifications, operational procedures and interviews. The conceptual model is represented using UML.

Collection and validation of data consisted of collection and scrutinizing of relevant documents (Tele2 internal documents, 2012). With the help of these documents and interviews, requirements for factors which should be considered during CR planning are stated. Each requirement was sent for approval to a representative change/release manager of the particular system. The missing information was supplemented by interviews. Consequently work was validated in the sense that every stakeholder was involved in study scrutinizing and finally the requirements and model was approved. Additionally, the conceptual model and findings were validated by cross-examination with the existing literature in order to check ability of the elaborated conceptual model to represent CR scheduling problems investigation by other researchers.

Requirements

Brainstorming and general requirements

During the first stage of analysis, concepts of ERP system change process and change planning were elicited. These are presented below in the brainstorm diagram (Fig. 1). This diagram was used as the basis for the next stage, where concepts were structured and supplemented with proper descriptions. All initial concepts are divided in five groups, namely, objectives of CR planning, constrains and factors affecting the planning process, variables influencing the decision-making as well as opportunities for improvements in the change management process.

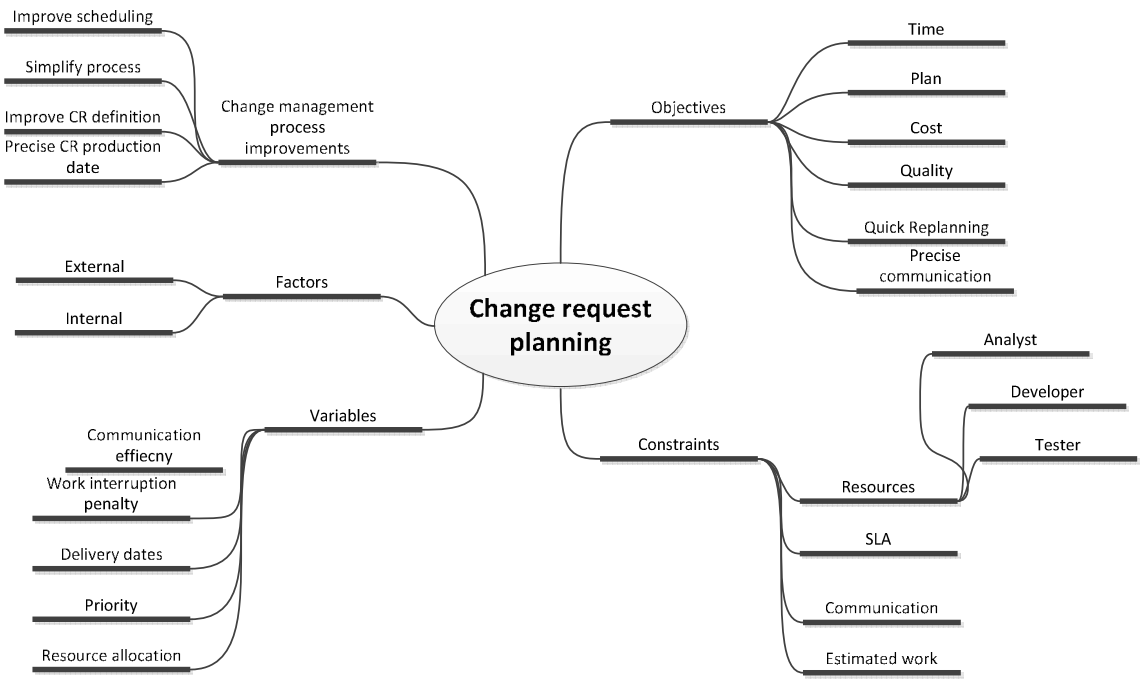


Fig. 1. **Brainstorm diagram.**

The main requirements towards the change management process and the CR planning algorithm are also identified:
1. The algorithm should support processing of the input CR list considering different parameters (factors) and yield the resource allocation and time plan. This is necessary to make the algorithm relevant to all possible factors which can happen in reality.
2. The algorithm should support re-planning at any point of time. This is necessary in case the priority or any of CR factors changes, new CRs are introduced or existing CR are removed from the list.
3. In re-planning, the algorithm should consider already performed work on the previously planned CRs. As the input CR list or resources can be changed during the time, such requirement is necessary.
4. The algorithm should support functionality to group certain CR from the pool; with intention to process implementation activity for the group of CRs. Regression testing could serve as example here: regression testing is performed for particular ERP software module or unit. If there are several CR related to one ERP software functional unit, the resource savings occur if all them are tested the same iteration.
5. The algorithm should support the following decision objectives: cost saving, plan accuracy with respect to delivery date, shortest implementation time.
6. Planned activities and resource allocation should be represented in a calendar time frame. This is necessary to be able to communicate the planning outcome to change managers or any stakeholders.
7. The algorithm should support human resource pool configuration. It should be possible to define efficiency, cost rate, availability of particular human resource.

Analysis of documents

Tele2 IT organization units have three main documents related to CR planning, namely, (1) the CR list collected from all requestors, (2) the list of available resources and (3) manual mapping of CR towards resources for building the change implementation plan.
The manual planning is driven by two main parameters. Requested production date is a primary parameter and priority is a secondary parameter. Requested production date is neither unique among CR nor mandatory, however it must be strictly observed. Priority is used to order CR with the same requested production date and to order CR without the requested production date. The market-driven requirement engineering processes (Daveport, 1988) have a strong focus on requirement prioritization (Carlshamre, 2002).
A sample CR list in given in Table 1 and a sample resource list is given in Table 2. In addition to production data and priority, each CR is characterized by effort required for change design, development and testing, changes description and functional module affected by the change. Each resource is characterized by its efficiency ranking, hourly rate and area of specialization.

Table 1

Example table of intake change request list

CR	Description	Design (h)	Development (h)	Tests (h)	Priority	Production date	Functional module
1.	Incorrect calculation formula	2	3	8	1	01.01.2012	Finance
2.	Incorrect description on invoice	3	1	12	2	01.02.2012	Invoicing
3.	Legal req. New VAT rate.	12	23	45	1	01.01.2012	Finance
4.	Adaptive change, new XML block in invoices.	12	45	0	1		Invoicing
5.	Low priority GUI fixes. Ordering in ComboBox.	2	2	12	5		GUI

Table 2

Example table of available resource list

ID	Type	Ranking (1..5)	Rate/hour (EUR/h)	Functional unit
1.	developer	1	25	CRM
2.	tester	5	50	Finance, CRM
3.	analyst	3	35	Reporting
4.	tester	2	25	Sales
5.	developer	1	60	Invoicing

Conceptual model

The conceptual model is elaborated according to the requirements gathered. These requirements allow to define the general approach to change implementation planning as shown in Fig. 2. The general approach shows that the planning algorithms receives CR list, resource list and other relevant factors as input data and the list of CR included in a current release and resource allocations are received as the planning outputs. The conceptual model formally defines these input and outputs used by the algorithm.

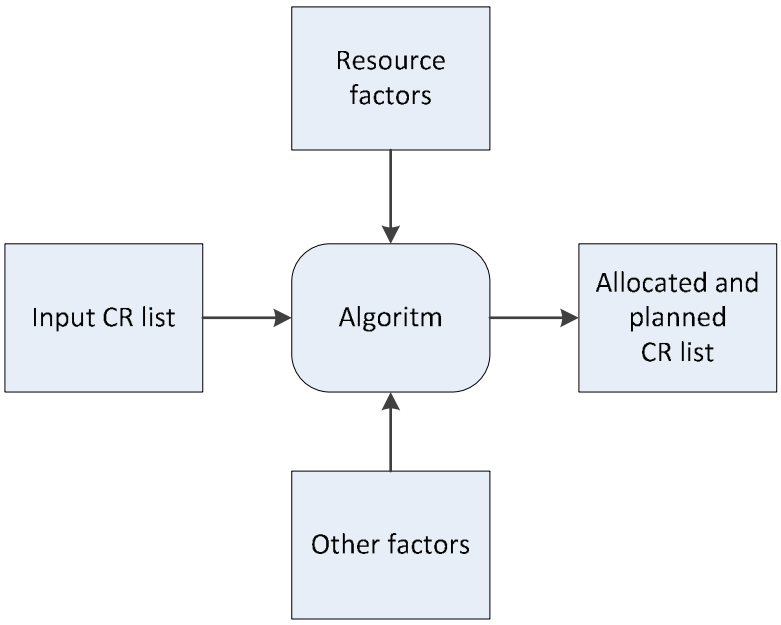


Fig. 2. **The high level description of ERP system change implementation planning.**

During research was identified limitation: each of CR should pass implementation time estimate analyses phase. Potentially, prerequisite for each CR should be provided resource requirements. There should be known how much time is needed from developer/designer/tester in order to fulfil request. The estimated hours might be provided based on average team efficiency (e.g. should not be build based on high performers in the team).

Description of concepts

The main concepts defining the CR planning problem are change requests, change request list, resource, resource list, objective, original requirement, functional software unit, change request plan and resource allocation plan.
The change request is a document containing a call for an adjustment of IT system. The main attributes of the change request are:

- Requested production date - calendar parameter. It stands for primary CR processing intake to model. The most production date critical CR should be allocated, if several CRs have the same date, a secondary priority parameter should be used.
- Priority - Subjective unique parameter. It stands for secondary CR processing intake to model. Model should allocate resource to CR consequently based on priority.

- Design (h) - time in hours required for the design of particular CR.
- Development (h) - time in hours required for the development of particular CR.
- Testing (h) - time in hours required for the testing of particular CR.
- Dependency – CR represented by this attribute can have implication or exclusion dependencies with other CRs. There are six types of dependencies (Li, 2007). However, only the implication dependency as most frequently occurring (Carlshamre et al., 2001) is considered in the paper and conceptual model.
- Functional unit - the functional unit of ERP software CR relates to. (E.g. Financial, GUI, Reporting and etc). It is used to group CRs and to plan one regression testing.

The resource is defined as human resource – the individual who makes up workforce of organization. The main attributes of the resource are:

- Type – defines type of the resource, for example, developer, designer, tester.
- Ranking - subjective parameter, which is represents efficiency of specialist, and might reduce estimated time for each activity.
- Rate/hour- actual costs per hour for particular human resource.
- Communication multiplier - if different types of resources are involved in CR implementation some of the development effort is spent on communication. This attribute represents the communication effort, which is only one recourse is involved. This attribute is also used to represent communications with vendors of ERP system.
- Work interruption penalty - resources are often shifted from one task to another to complete new high priority requests. If work on current CR is interrupted, there is a resumption delay when resource returns back this CR. The work interruption penalty attribute accounts for this delay.
- Functional unit - specialization in functional units of ERP system which can be addressed to this resource. Maps to functional unit of CR.
- Availability - planned vacation days, sickness and etc.

Standard regression test estimates for functional unit - Standardized (agreed) time frame which is required to cover functional unit of ERP system with regression tests.

The objective attribute represents the aim of CR planning. The following objectives can be considered: cost saving, time accuracy, the shortest implementation time. Each of these modes can be selected running planning or re-planning.

Associations among concepts

Based on the finding in the research there was created conceptual model, considering found concepts and their attributes. Conceptual model was build using UML class diagram notation (Fig. 3) (Larman, 2005). The concept attributes are not represented on diagram, and are provide in Table 3.

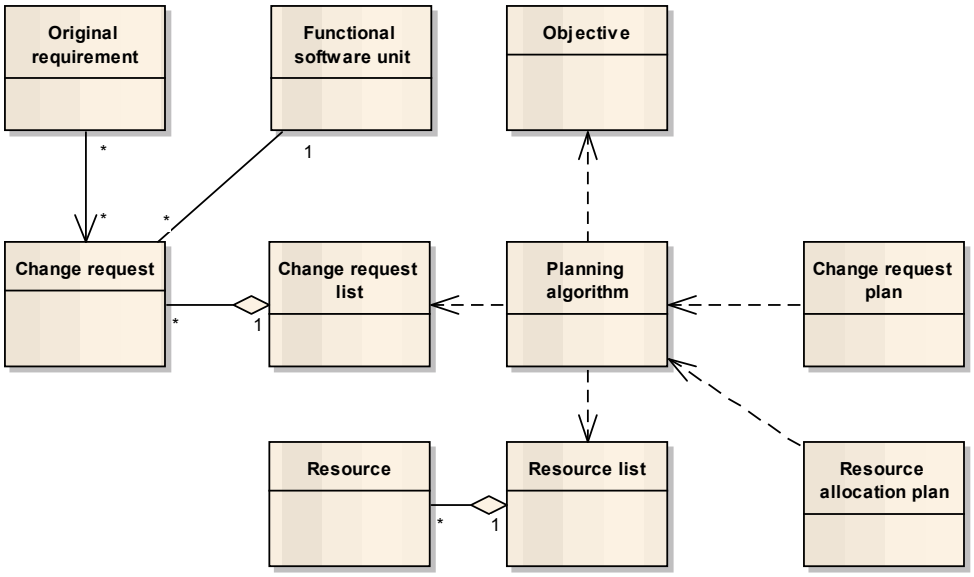


Fig. 3. **Conceptual model.**

Validation

The conceptual model, concepts and their attributes constructed during research were validated against several existing studies in this field. Four papers (Li et al., 2007; Reboucas et al., 2006; Sauvé1 et al., 2006; Carlshamre

et al., 2001) dealing with CR implementation planning were selected for validation. Table 3 marks concepts and attributes from the conceptual model also discussed in the related papers. The validation result shows that the conceptual model is able to represent CR planning factors considered in the related studies and is more comprehensive than existing CR implementation planning models.

Table 3

Concept and their attributes validation

Concept and attributes	Li. C. et al., 2007	Reboucas R. et al., 2006	Sauvé J. et al., 2006	Carlshamre P. et al., 2001
Change Request	x	x	x	X
Requested_production_date	x	x	x	
Priority	x	x	x	x
Required_Design_Time	x	x		
Required_Development_Time	x	x		
Required_Test_Time	x	x		
Dependency on CR	x	x	x	x
Functional_Unit_ID				
Change request List		x		
Resource	x	x	x	x
Ranking	x		x	
Rate_hours	x	x	x	
Communication_multiplier				
Work_interruption_penalty				
Functional_Unit_ID				
ResourceList		x		
Functional units	x		x	
ID	x		x	
Name	x		x	
Required_regression_tests_hours				
Decision directions(cost saving, time saving, precise communication)	x	x	x	x
ID	x	x	x	
Name	x	x	x	
Change Request Plan		x	x	x
ID			x	x
Ordering			x	x
Planned production date			x	
Resource Occupation Plan		x	x	
ID			x	
Change_Request_ID			x	
Hours			x	
Occupation_Date			x	

The cross-evaluation also revealed several important and challenging aspect of CR implementation planning. Importance of CR planning as analysed in this paper is confirmed by Par Carlshamre (Carlshamre et al., 2001): “Release planning is a crucial activity in market–driven software development, because it decided what should be delivered and when. Release planning is matter of prioritization the requirements and selecting a number of top priority requirements depending on the available resources and delivery date.” Related research also

emphasizes importance of the "priority" attribute: "priority of requirements is a major determinant in incremental planning" (Carlshamre et al., 2001).
The further study by Li C. et al. (2007) focus on such attributes as priority and efforts (costs/time): "lists of such variables, including: importance or business value, stakeholder preference, cost of development, requirement quality, development risk" (Li et al., 2007). The study also confirmed on-time-delivery as driver for the model, where the same attribute was selected as major in this paper. However, their model does not consider the decision objective of cost saving, which is one of the main decision-making objectives in release planning.
Re-planning of CR after changes in the CR list is also investigated by Jacques Sauvé et al. (2006): „we identified the main challenges in IT change management as planning and scheduling of changes; high number of emergency changes". ITIL's change management process recommendations ranked the first 3 most important change management challenges as being (Sauvé et al., 2006):

1. Scheduling/planning changes (with 47 points out of a maximum 55, or over 85%);
2. High number of emergency changes (43 points or 78%);
3. Request for change scope ill-definition (40 points or 72%).

The first and second points are covered within this paper by providing the re-planning feature.

Discussion and conclusion

According to the web survey (Sauve et al., 2006), the main challenges in IT change management are 1) planning/scheduling changes, 2) high number of emergency changes and 3) ill-definition or wrong scoping of request for changes. In this paper we addressed the problem of planning and scheduling change requests, taking in account numerous relevant factors.
Analysis of the change implementation process at Tele2 IT division was performed and described in this paper. Main CR planning concepts and their attributes as well as the planning approach were identified.
These are represented in the CR planning conceptual model. The model f was validated against the existing CR planning methods reported in literature, and synergies and gaps were identified. The conceptual model serves as the basis for further elaboration of the CR planning algorithm and decision support system to be used by change managers.
The main contribution of this paper is conceptualization of the CR planning problem. Design and implementation of the planning support tool is the subject of the next phase of our research.
The current conceptual model is limited by an assumption that all parameters characterizing CR and resources are constant throughout the planning horizon. However, it is common in different IT companies that some of these parameters might change. For example, a resource changes its original specialization during implementation in order to meet the deadline.

Acknowledgements

The authors would like to thank the Tele2 change/release managers that participated in the research questionnaire.

References

Bailor and Coreen, 2006. "For CRM, ERP, and SCM, SAP Leads the Way".
Boehm, B.W., Parikh, G., 1988. Techniques of Program and System Maintenance, QED Information Sciences, Inc. pp. 51.
Carlshamre, P., 2002. Release Planning in Market-Driven Software Product Development: Provoking an Understanding. Requirements Engineering 7(3). pp. 139–151.
Carlshamre, P., Sandahl, K., Lindvall, M., Regnell, B., Natt och Dag, J., 2001. An industrial survey of requirements interdependencies in software release planning. In: Proceedings of the 5th IEEE international symposium on requirements engineering, pp. 84–91.
Carlshamre, P. and Regnell, B., 2000. Requirements Lifecycle Management and Release Planning in Market-Driven Requirements Engineering Processes. International Workshop on the Requirements Engineering Process: Innovative Techniques, Models, and Tools to support the RE Process, 6th-8th of September, Greenwich, UK, the DEXA Conference.
Chuck, C.H., Law, Charlie, C., Chen, Bruce, J.P., Wu, 2010. Managing the full ERP life-cycle: Considerations of maintenance and support requirements and IT governance practice as integral elements of the formula for successful ERP adoption, Journal Computers in Industry 61 (3). pp. 297-308.
Daniel, L. and Moody, 2005. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions, Gerstner Laboratory, Czech Technical University, Prague 13000, Czech Republic.
Daveport, T., 1988. Putting the enterprise into the enterprise systems, Harvard Business Review 76 (4). pp. 121–131.
Firesmith, D., 2004. Prioritizing Requirements. Journal of Object Technology 3(8). pp. 35–47.

Florae, W.A., 1992. Software quality measurement: A framework for counting problems and defects, Technical Report CMUX.EI-92-TR-22, Software Engineering Institute, Carnegie-Mellon University, Pittsburgh, PA Sprott D, 2000.

Greer, D. and Ruhe, G., 2004. Software release planning: an evolutionary and iterative approach. Information and Software Technology 46. pp. 243–253.

Larman, C., 2005. Applying UML and patterns: an introduction to Object-Oriented Analyses and Design and Iterative Development. - 3rd ed. pp. 134-159.

Li, C., van den Akker, J.M., Brinkkemper, S., and Diepen, G., 2007. Integrated Requirement Selection and Scheduling for the Release Planning of a Software Product

Lientz, B. and Swanson, E., 1981, Problems in application software maintenance, Communications of the ACM 24 (11). pp. 763–769.

Reboucas, R., Sauve, J., Moura, A., Bartolini, C., Trastour, D., 2006. "A decision support tool to optimize scheduling of it changes," in *Proceedings of IFIP/IEEE Network Operations and Management Symposium, Munich, Germany*.

Sauvé, J. Rebouças, R., Moura, A., Bartolini, C., Boulmakoul, A., Trastour, D., 2006. Large Scale Management of Distributed Systems Lecture Notes in Computer Science Volume 4269. Business-Driven Decision Support for Change Management: Planning and Scheduling of Changes. pp. 173-184

Sprott, D, 2000. Componentizing the enterprise application packages. Communications of the ACM, 43(4). pp. 63–69.

Tele2 internal documents., 2012.

Weerd, I., van de Brinkkemper, S., Nieuwenhuis, R., Versendaal, J.M., Bijlsma, A., 2006. Towards a Reference Framework for Software Product Management. In: Glinz, M., Lutz, R.R. (eds.) 14th IEEE International Requirements Engineering Conference, Minneapolis/St. Paul, Minnesota, pp. 319–322. IEEE Computer Society, Washington.

Yang and Xin-She, 2010. Engineering Optimization: An Introduction with Metaheuristic Applications, Wiley Hoboken, NJ, USA.

Zia, L., 2006. Optimizing Change Request Scheduling in IT Service Management, IEEE International Conference on Services Computing.

Appendix

Concepts and their attributes

Concept	Attributes
Change Request	ID Number Title Text Requested_production_date Date Priority 0..1 Number Required_Design_Time Date Required_Development_Time Date Required_Test_Time Date Dependecy_on_CR Number Functional_Unit_ID Func_Unit
Change Request List	ID Number
Resource	ID Number Type Text Ranking Number Rate_hours Number Communication_multiplier Date Work_interruption_penalty Date Functional_Unit_ID Func_Unit
Resource List	ID Number
Functional units	ID Number Name String Required_regression_tests_hours Number
Decision directions	ID Number Name Dec_directions
Change Request Plan	ID Number Ordering Number Planned production date Date
Resource Occupation Plan	ID Number Change_Request_ID Number Hours Number Occupation_Date Date

Coloured Petri Nets based simulation scheme for adaptive bandwidth management

Julija Asmuss, Gunars Lauks

Institute of Telecommunications, Riga Technical University, Azenes str. 12, LV-1048, Riga, Latvia
julija.asmuss@rtu.lv, gunars.lauks@rtu.lv

Abstract: *We consider the problem of resource allocation for multiple classes of traffic in a substrate network with DaVinci architecture. According to DaVinci (Dynamically Adaptive Virtual Networks for a Customized Internet) approach a single physical network can support multiple traffic classes with different performance objectives by means of multiple virtual networks constructed over the physical one. In this context the problem of bandwidth allocation is a conditional maximization problem for the aggregate performance of all virtual networks. We describe Coloured Petri Nets (CPN) based models and present CPN based schemes for simulation of dynamically adaptive bandwidth allocation mechanisms proposed to support multiple traffics in a substrate network. The simulation study focuses on two traffic types: delay sensitive and throughput sensitive. Obtained simulation results demonstrate the achievement of dynamically adaptive bandwidth management strategies.*

Keywords: Bandwidth Allocation Problem, Coloured Petri Nets, Simulation.

Introduction

The current Internet carries different types of services (voice, video, music, web pages, e-mail). Some services need low delay mechanisms while other services need high throughput mechanisms, and their requirements may conflict with each other. Network virtualization principles (Anderson et al., 2005) can be used for building experimental platforms that run multiple virtual networks. From network point of view network virtualization divides a network into a set of virtual networks. These virtual networks give us the opportunity to classify and separate traffic. Each virtual network is logically separated and can be customized for a particular traffic class. Resources offered by a substrate network are shared between all virtual networks. Finding the proper bandwidth allocation to virtual networks is one of the key problems of network virtualization (Szeto et al., 2003; Zhu et al., 2006; Haider et al., 2009; Dramitinos, 2009; Zhang et al., 2009; Zhou et al., 2010).
DaVinci approach (Dynamically Adaptive Virtual Networks for a Customized Internet) describes a technique of network virtualization, according to which all virtual networks are constructed over the physical substrate network by subdividing each physical node and each physical link into multiple virtual nodes and virtual links (He et al., 2008). We consider the problem of bandwidth resource management in a substrate network on the basis of DaVinci architecture. In this context it is a maximization problem for the aggregate utility of all virtual networks (Lin et al., 2006), which effective solution depends on the design of dynamically adaptive bandwidth allocation protocols. When different types of traffic coexist over the same substrate network, each virtual network could control a subset of resources at each node and link. At a smaller timescale each virtual network maximizes its own utility. The question is whether optimization of virtual networks together with the bandwidth share adaptation scheme performed by the substrate network actually maximizes the aggregate utility.

Bandwidth allocation problem

The DaVinci architecture (He et al., 2008) allows us to describe how a single substrate network can support multiple traffic classes, each with a different performance objective. We describe the topology of a substrate network by a graph $G_S = \{V_S, E_S\}$, given by a set V_s of physical nodes (or vertices) and a set E_S of physical links (or edges). We suppose that links of E_S are with finite capacities C_l (links are denoted by l: $l \in E_S$). Correspondingly to $G_S = \{V_S, E_S\}$ we consider DaVinci model with virtual networks, indexed by k, where $k = 1, 2, ..., N$. According to the DaVinci approach, each traffic class is carried on its own virtual network with customized traffic-management protocols. The substrate runs schedulers that arbitrate access to the shared node and link resources, to give each virtual network the illusion that it runs on a dedicated physical infrastructure. Let the key notations for virtual network k, $k = 1, 2, ..., N$, be the following:

- $\mathbf{y}^{(k)}$ – bandwidth of virtual network k,
- $\mathbf{z}^{(k)}$ – path rates for virtual network k,
- $\boldsymbol{\lambda}^{(k)}$ – satisfaction level degree of virtual network k,
- $U^{(k)}$ – performance objective for virtual network k.

Bandwidth values $\mathbf{y}^{(k)} = (y_l^{(k)})_{l \in E_s}$ for each substrate link $l \in E_S$ are assigned by the substrate network, taking into account such local information as current satisfaction indicators and performance objectives. The substrate network periodically reassigns bandwidth shares $\mathbf{y}^{(k)}$ for each substrate link between its virtual links. Thus, values $\boldsymbol{\lambda}^{(k)} = (\lambda_l^{(k)})_{l \in E_s}$ and $U^{(k)}$ are periodically updated by the substrate network and used to compute virtual link capacity $\mathbf{y}^{(k)}$.

The objective function $U^{(k)}$ depends on both virtual link rates $\mathbf{z}^{(k)}$ and virtual link capacity $\mathbf{y}^{(k)}$. The objective is subject to a capacity constraint and possibly other constraints described in terms of $g^{(k)}(\mathbf{z}^{(k)})$. The goal of the substrate network is to optimize the aggregate utility of all virtual networks

$$\sum_{k=1}^{N} U^{(k)}(\mathbf{z}^{(k)}, \mathbf{y}^{(k)})$$

under constraints

$$\sum_{k=1}^{N} \mathbf{y}^{(k)} \le \mathbf{C}, \quad \mathbf{H}^{(k)}\mathbf{z}^{(k)} \le \mathbf{y}^{(k)}, \quad g^{(k)}(\mathbf{z}^{(k)}) \le 0, \quad \mathbf{z}^{(k)} \ge \mathbf{0}, \quad k = 1, 2, ..., N.$$

The capacity constraint requires the link load

$$\mathbf{r}^{(k)} = \mathbf{H}^{(k)}\mathbf{z}^{(k)}$$

to be no more than the allocated bandwidth. To compute the link load we use routing indexes

$$H_{lj}^{(k)i} = \begin{cases} 1, & \text{if path } j \text{ of source } i \text{ in virtual network } k \text{ uses link } l, \\ 0, & \text{otherwise,} \end{cases}$$

and path rates $z_j^{(k)i}$ that determine for source i the amount of traffic directed over path j.

An optimization scheme follows directly from DaVinci principles. First, the substrate network determines how satisfied each virtual network is with its allocated bandwidth. Satisfaction level degree $\lambda_l^{(k)}$ (for link l of virtual network k) is an indicator that a virtual network may want more resources. Next, the substrate network determines how much bandwidth virtual network k should have on link l: the substrate network increases value $y_l^{(k)}$ proportional to the satisfaction level $\lambda_l^{(k)}$ on link l.

Coloured Petri Nets based models

The concept of Coloured Petri Nets (Jensen, 1992–1997) is an extended version of classical Petri Nets. In addition to places, transitions and tokens, the concept of types or colour sets is included. This concept enables to involve information (simple or complex) into the tokens and allows the use of tokens that carry data values and can hence be distinguished from each other. Each token could be attached with a colour, indicating the identity of the token. Moreover, each place and each transition has attached a set of colours. A transition can fire with respect to each of its colours. By firing a transition, tokens are removed from the input places and added to the output places in the same way as that in original Petri Nets, except that a functional dependency is specified between the colour of the transition firing and the colours of the involved tokens.

A Coloured Petri Net is a tuple $CPN = (P, T, F, \Sigma, W, C, G, H, I)$ satisfying the following requirements:

- P is a finite set of places;
- T is a finite set of transitions, $P \bigcap T = \varnothing$, $P \bigcup T \ne \varnothing$;
- F is a set of directed arcs, $F \subset (P \times T) \bigcup (T \times P)$;
- Σ is a finite set of types (colour sets), $\Sigma \ne \varnothing$;
- W is a finite set of typed variables, $Type(w) \in \Sigma$ for all $w \in W$, where $Type : W \to \Sigma$ is a type function assigning types (colour sets) to variables;
- $C : P \to 2^{\Sigma}$ is a colour function assigning colour sets to each place, $C(p) \subset \Sigma$ for all $p \in P$;
- $G : T \to EXPR(W)$ is a guard function assigning a guard $G(t)$ to each transition $t \in T$ (we omit the explanations on the set of expressions $EXPR(W)$), such that $Type(G(t)) = Bool$ for all $t \in T$, where $Bool = \{true, false\}$;
- $H : F \to EXPR(W)$ is an arc expression function assigning an expression $H(f) \in F$ to each arc $f \in F$, $Type(H(p,t)) \in C(p)$ for all arcs $(p,t) \in F$ and $Type(H(t,p)) \in C(p)$ for all arcs $(t,p) \in F$;

- I is an initialisation function assigning an initial marking to each place, an initial marking can be defined as a multi-set $M_0 \in N^{PLACE}$, where $PLACE = \{(p,c) : p \in P, c \in C(p)\}$.

Triple (P,T,F) constitutes the net structure, pair (Σ,W) describes types and variables and tuple (C,G,H,I) defines the net inscriptions. Here we omit the explanations on marking iterations and do not discuss how transitions change the marking of places. Due to this iteration scheme, CPN is one of efficient mathematical modelling languages for the description of discrete event systems. CPN combines a well-developed mathematical theory with an excellent graphical representation. This combination is the main reason for the great success of CPN in modeling of the dynamic behaviour of systems (Jensen, 1992–1997; Kristensen et al., 1998; Jensen et al., 2007; Gehlo et al., 2010).

Coloured Petri Nets, proposed by Kurt Jensen, have been developed by the CPN group at Aarhus University, Denmark since 1979. The first version was a part of the PhD Thesis of Kurt Jensen and was published in 1981. The CPN group has developed and distributed industrial-strength computer tools such as Design/CPN in 1990 and CPN Tools in 2003. Our simulation scheme is based on Coloured Petri Nets Tools (Ratzer et al., 2003; Jensen et al., 2007). CPN Tools is a discrete event modeling computer tool for CPN models supporting interactive and automatic simulations, state spaces and performance analysis, and combining Coloured Petri Nets and the functional programming language CPN ML, which is based on Standard ML.

Colours can be effectively used for modelling virtual networks accordingly to the DaVinci architecture. A CPN model of a substrate network describes the states of each virtual network of the system and the events (transitions) that can cause the system to change state. By making simulations of the network CPN model with CPN Tools it is possible to investigate different scenarios and explore the behaviours of the system, to use simulation-based performance analysis for decision making and adaptation processes.

Simulation scheme

The simulation study focuses on two traffic types: delay sensitive and throughput sensitive. Accordingly to DaVinci principles we consider two virtual nets for two types of traffic denoted by *A* and *B*. Up to now we experiment with two nodes topology (Fig. 1) and use the following notations: G_A, G_B – packet generators; D_A, D_B – destination nodes.

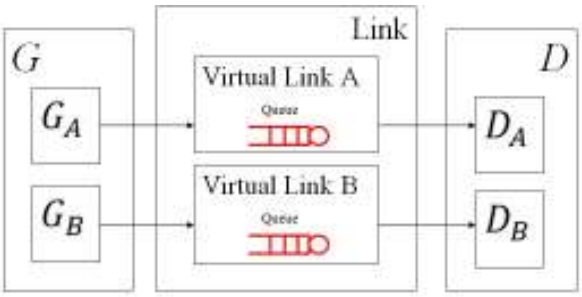


Fig. 1. **Simulation scheme for two nodes topology.**

Colours *A* and *B* are effectively used for modelling and simulating such system. The colours and variables associated with the tokens which represent packets are defined as follows:

```
colset PacketType = with A | B;
colset PacketSize =int;
colset Packet = record packetType : PacketType * packetSize : PacketSize * AT:int;
colset Packets = list Packet;
var packet: Packet;
var packets: Packets.
```

We also use colours *A* and *B* as identifiers for the corresponding virtual links. It is the number of tokens and the token colours which represent the state of a virtual link:

```
colset VirtualLinkType = with A | B;
colset VirtualLinkStatus = with VL_Status timed;
colset VirtualLinkBandwidth = int;
colset VirtualLink = record VLtype : VirtualLinkType * VLstatus : VirtualLinkStatus * VLbandwidth :
VirtualLinkBandwidth;
colset VirtualLinkxPacket = product VirtualLink * Packet timed.
```

In addition to token colours we use time stamps. As usual in timed CPN models, we use Model Time Units (MTU) and define additional time variables:

colset INT = int;
var proctime : INT.

The CPN Tools model (Fig. 2) involves transmission modules and bandwidth shares adaptation modules:

- Arrivals;
- Virtual Link;
- To Bandwidth Adaptation;
- From Bandwidth Adaptation.

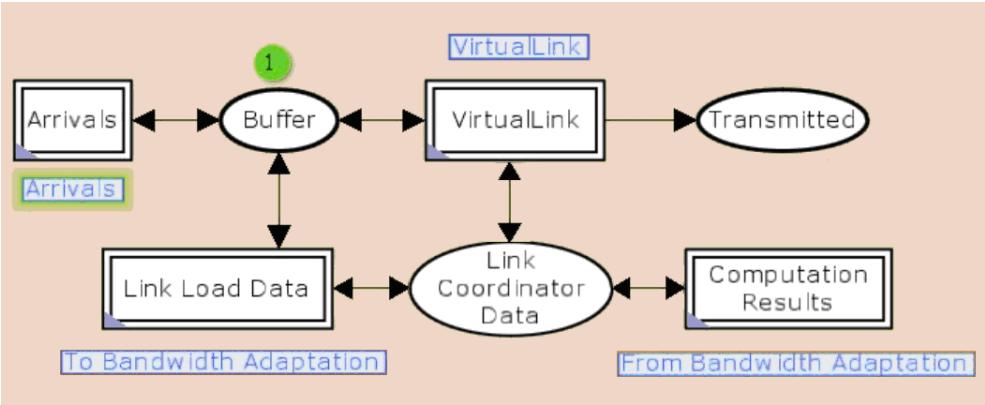


Fig. 2. **Bandwidth adaptation simulation scheme using CPN Tools.**

Packets are generated by the traffic generators (Arrivals) and stored in the FIFO queues (Buffer). The goal is to route packets to the output port (Transmitted). Packets cross the FIFO queue and are transmitted under the condition that the transmission link is free. This condition fulfilled if the place Idle (Fig. 3) is marked. If this place is empty, then the transmission link is busy. Technically token VL_Status is used to organize the queue and packets transmission as it is shown in Fig. 3. The transmission time depends on the size of a packet and the bandwidth of the corresponding virtual link and is calculated for each packet transmission. Module Link Load Data (Fig. 2) assigns a bandwidth value for each virtual link. Two traffic FIFO queues and two virtual links are separated due to colours *A* and *B* and two flows of packets are parallel controlled and analysed.

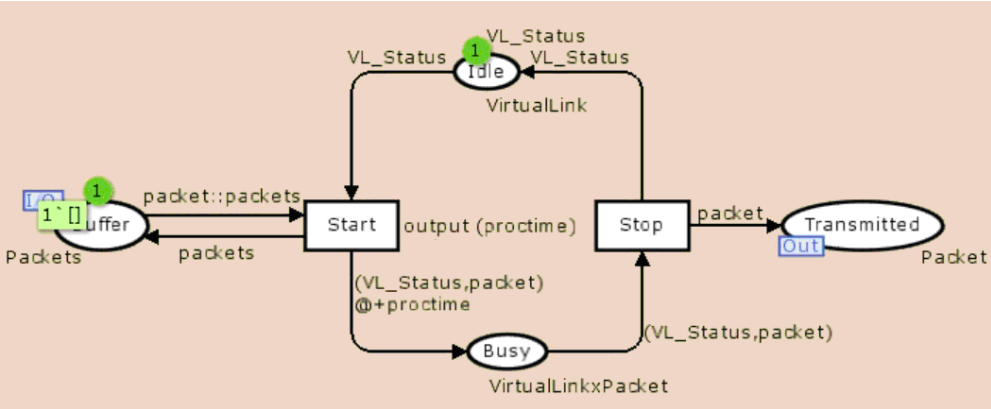


Fig. 3. **Virtual link simulation scheme using CPN Tools.**

Bandwidth allocation adaptation modules are designed to update the virtual link resources allocation for dynamically changing traffics. Special predicate and observation functions (fun pred, fun obs) and data collection monitors (Buffer Delay, Buffer Length, Virtual Link Utilization) are included for monitoring the performance of the system. A monitoring mechanism is used not only to control, but also to modify a simulation of the net. It is done by reassigning bandwidth shares between virtual links *A* and *B*. A decision making system is based on data collection monitors that allow to calculate the system performance measures such as the delay in each queue, the length of each queue, the utilization of each link. The criteria for decision making depend on the aggregate objective function for both virtual links.

In our experiments the initial resource allocation to virtual links is uniform, we set the capacity of the physical network as 100 Mbps. Packets *A* and *B* are generated with exponentially distributed arrival time and uniform distributed size. An important issue is the frequency of adaptation. Bandwidth resources are reassigned every 10000 MTU. By changing traffic parameters every 50000 MTU we observe the adaptation process and obtain good adaptation results after 2-3 adaptation iterations.

We experiment with delay sensitive traffic (the objective is to minimize the average delay) and throughput sensitive traffic (the objective is to maximize the average link rate) as *A* and *B* correspondingly, as well as with two delay sensitive traffic classes and with two throughput sensitive traffic classes. Our simulation results clearly show that the adaptive bandwidth allocation mechanism can dynamically and efficiently react to traffic changes in both cases: when traffic classes are with different performance objectives or with the same one.

Conclusion

In this paper we present the design and experimental evaluation of a simulation scheme for an adaptive bandwidth allocation mechanism, which is realized for two nodes topology. This simulation scheme is based on Coloured Petri Nets and realized by using CPN Tools. The effectiveness of our scheme is evaluated within simulation experiments with two types of traffic: delay sensitive and throughput sensitive.

We see several possible extensions of this work. Firstly, our future work will focus on extension of the proposed simulation scheme in order to generalize the network topology. Instead of two nodes topology we plan to consider Access-Core topology, Abilene topology, etc. We should extend the model from local link level to global network level by involving a global coordinating mechanism. It is clear that the virtual link bandwidth assignment on one physical link would be dependent on the virtual links bandwidth assignments on another physical links. Of course, the model mentioned above requires more complicated bandwidth allocation scheme. Another future extension will generalize the proposed simulation scheme from two traffic classes to multiple traffic classes by involving additional colours in the CPN model. At the same time we will enhance our model with more functions in order to modify and improve monitoring and decision making systems and, as a result, to optimize adaptive bandwidth management.

References

Anderson, T., Peterson, L., Shenker, S. and Turner, J., 2005. Overcoming the Internet Impasse through Virtualization. *IEEE Computer*, 38 (4), pp.34-41.

Dramitinos, M. L., 2009. A Bandwidth Allocation Mechanism for 4G. *Proceedings of European Wireless Technology Conference*, Rome, pp.96-99.

Gehlo, V. and Nigro, C., 2010. An Introduction to System Modeling and Simulation with Coloured Petri Nets. *Proceedings of Winter Simulation Conference,* Baltimore, Maryland, USA, pp.104-118.

Haider, A., Potter, R. and Nakao, A., 2009. Challenges in Resource Allocation in Network Virtualization. *Proceedings of the 20th ITC Specialist Seminar*, Hoi An, Vietnam. Available at: www.itcspecialistseminar.com/paper/itcss09_Haider.pdf, 20.12.2012.

He, J., Zhang-Shen, R., Li, Y., Lee, C.-Y., Rexford, J. and Chiang, M., 2008. DaVinci: Dynamically Adaptive Virtual Networks for a Customized Internet. *Proceedings of ACM CoNEXT Conference,* Article No. 15, ACM, New York, USA. Available at: www.cs.princeton.edu/~jrex/papers/davinci.pdf, 20.12.2012.

Jensen, K., 1992–1997. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use.* Volumes 1-3. Springer-Verlag.

Jensen, K., Kristensen, L. and Wells, L., 2007. Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems. *International Journal on Software Tools for Technology Transfer*, 9, pp.213-254.

Jensen, K. and Kristensen, L., 2009. *Coloured Petri Nets. Modelling and Validation of Concurrent Systems*. Springer-Verlag.

Kristensen, L., Christensen, S. and Jensen, K., 1998. The Practitioner's Guide to Coloured Petri Nets. *International Journal on Software Tools for Technology Transfer*, 2, pp.98-132.

Lin, X. and Shroff, N., 2006. Utility Maximization for Communication Networks with Multi-Path Routing. *IEEE Trans. Automatic Control,* 51, pp.766-781.

Ratzer, A.V., Wells, L., Larsen, H.M., Laursen, M., Qvortrup, J.F., Stissing, M.S., Westergaard, M., Christensen, S. and Jensen, K., 2003. CPN Tools for Editing, Simulating and Analysing Coloured Petri Net. *Lecture Notes in Computer Science*, 2679, pp.450-462.

Szeto, W., Iraqi, Y. and Boutaba, R., 2003. A Multi-Commodity Flow Based Approach to Virtual Network Resource Allocation. *Proceedings of the Global Communications Conference IEEE GlobeCom,* San Francisco, CA, USA, pp.3004-3008.

Zhang, Y., Wang, C. and Gao, Y., 2009. A Virtualized Network Architecture for Improving QoS. *International Journal on Information Engineering and Electronic Business*, 1, pp.17-24.

Zhou, Y., Li, Y., Sun, G., Jin, D., Su, L. and Zeng, L., 2010. Game Theory Based Bandwidth Allocation Scheme for Network Virtualization. *Proceedings of the Global Communications Conference IEEE GlobeCom2010*, Miami, Florida, USA. Available at: http://ieeexplore.ieee.org, 20.12.2012.

Zhu, Y. and Ammar, M., 2006. Algorithms for Assigning Substrate Network Resources to Virtual Network Components. *Proceedings of the 25th IEEE International Conference on Computer Communications InfoCom2006*, Barcelona, Spain, pp.1-12.

Knowledge modeling for ontology-based multiatrribute classification system

Anna Galinina, Arkady Borisov

Department of Modelling and Simulation, Riga Technical University, 1 Kalku Street, Riga, LV-1658, Latvia
asr-89@mail.ru, arkadijs.borisovs@rtu.lv

Abstract: *Developers of the early knowledge-based systems expected that their systems would be ideally suited for reuse (Hayes-Roth, Waterman, & Lenat, 1983). These systems consisted of two parts: a knowledge base containing rules about the problem domain and an inference engine that used the knowledge base to provide solutions. The hope was that this approach could be applied to a variety of different domain tasks. Unfortunately, the overhead cost of constructing knowledge bases for new domains was often too high: the knowledge-acquisition "bottleneck" problem arose because it is difficult to describe and encode domain knowledge in the low-level terms required by the inference engine. The purpose of this paper is ontology bifurcation into domain ontology (a set of concepts and relations independent of any classification method) and method (application) ontology (set of concepts and relations) devoted to decision tree classification method.*

Keywords: method ontology, ontology construction, multi-attribute classification.

Introduction

Ontologies can be applied successfully in solving the problems of classification and clustering. In all these cases, the used ontology is a set of meaningful concepts, defining the subject area. Their application allows avoiding loss of computation time for the analysis of concepts not included in the subject area, and in the case of the classification problem - not learning sometimes very costly classifier in the training set, as represented by the classifier composed ontology (Ben-David, 2010). On the other hand, the quality of the solution of these problems is very dependent on the quality and completeness of the ontology composed. In classification systems and information systems ontology can be used to:

- Share a common understanding of information structure among people or software agents;
- Reuse the domain knowledge;
- Make assumptions explicit in the subject area;
- Separate the domain knowledge from the operational knowledge;
- Analyze the domain knowledge.

Decision tree classifiers

The development of information technology, in particular, progress in the methods of collection, storage and processing of data has enabled many organizations to collect large amounts of data to be analyzed. The amount of data is so large that the capacity of experts is not enough, which has led to the demand for automatic methods of research (analysis) of data, which every year is increasing. Decision trees is one of the methods of automatic data analysis (Rokach, 2002).

Decision trees are a way of representing hierarchical rules, sequential structure, where each object has a single node, giving the solution. The rules mean a logical construction, presented in the form of "if ... then ...".

Decision trees can be used widely, but all the problems solved by this unit can be grouped into the following three categories:

- Data Description: Decision trees can store information about the data in a compact form, instead of them, we can store the decision tree, which contains an accurate description of the objects.
- Classification: Decision trees are doing a fine job with the tasks of classification, assigning an object to one of the previously known classes. The target variable must have discrete values.
- Regression: If the target variable has continuous values, decision trees allow you to set the target variable dependence on the independent (input) variables. For example, this class contains the problem of numerical prediction (predicting the target variable values) (Polaka, 2010).

Ontology engineering

The separation of knowledge specifications into two components (canonical form and ontology) introduced by Gruber is not very convenient, because the same knowledge may be described redundantly. Modern languages for defining ontologies allow combining these forms of specifications together. So now the ontology is defined as any description of declarative knowledge, made using a formal language and provided with a classification specifiable knowledge, allowing a person to easily perceive it. The canonical form is not necessarily using the language of predicate logic. For example, you can use the so-called algebraic approach to describe knowledge, in

which the facts are presented in the form of terms, and the different relationships between the facts - in the form of restrictions, expressed in the form of axioms of equivalence. But any such description should include representation of declarative knowledge as a hierarchy of objects (classes), only in this case this description can be considered as ontology.

Often the term “ontology” is used only for knowledge specification about the world. It is conceptualization to describe structure of the world regardless of any engineering problem. This type of conceptualization is typical for philosophers. In philosophy, the term "ontology" is used in the sense of specification of knowledge about the world. Programmers, however, have a different kind of problem: they use conceptualization in order to construct a model of the problem being solved. Thus, philosophers and programmers have different objectives when conducting conceptualization: the first aim to describe the properties of reality and the formal model of a specific task (Prabowo et al., 2002).

An ontology defines the common words and concepts (meanings) used to describe and represent an area of knowledge, and so standardizes the meanings. Ontologies are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area or area of knowledge, like medicine, counterterrorism, imagery, automobile repair, etc.). Ontologies include computer usable definitions of basic concepts in the domain and the relationships among them. They encode knowledge of a domain and also knowledge that spans domains. So, they make that knowledge reusable.

An ontology includes the following:

- Classes (general things) in the many domains of interest;
- Instances (particular things);
- Relationships among those things;
- Properties (and property values) of those things;
- Functions of and processes involving those things;
- Constraints on and rules involving those things (Guarino, 1998).

Ontological engineering denotes a set of design principles, development processes and activities, supporting technologies, and systematic methodologies that facilitate ontology development and use throughout its life-cycle design, implementation, evaluation, validation, maintenance, deployment, mapping, integration, sharing, and reuse. Ontological engineering provides a design rationale for the development of knowledge bases and enables systematization of knowledge about the world of interest and accumulation of knowledge (Niaraki, 2009).

Knowledge engineering for an intelligent system should always include ontological engineering, which implies using specific development tools and methodologies.

Related works

The term “ontology” was proposed by Thomas Gruber in his work (Gruber, 1991) where the various aspects of interactions between intelligent systems and human were considered. This interaction may be difficult, but it is always of the same type. Therefore all knowledge of a program is static. Intelligent system in this sense is more versatile - it has the knowledge of what to do in the program and it can vary. Knowledge that is incorporated into computer programs can be divided into two classes:

- Procedural knowledge, i.e. knowledge about what to do in each situation;
- Knowledge of the problem world or declarative knowledge, i.e. payments, transactions, accounts, etc. Without this knowledge, it is obvious, the program will not function, and an algorithm cannot be constructed of the software system.

If the inside of the intellectual system of knowledge about the world can be coded without a framework, to share this knowledge with other intelligent system one must provide a description of this knowledge. This description must be sufficiently formal to be understandable to another system, and must be known to the language of the description. In addition, the description should be clear. Gruber suggested describing this knowledge in two ways:

- In the canonical form, which is a description of the knowledge in the language of predicate logic (e.g., in the form of facts language Prolog);
- In the form of ontology that is a set of classes connected by a generalization relationship (the inverse relation to inheritance relationships).

So, ontology according to Gruber is a description of declarative knowledge, described in the form of classes with a relation of hierarchy between them. This description, intended for human consumption, is complemented by a description in canonical form, which is designed for reading by machines. Drafting description of declarative knowledge is usually a lot of work and requires specific skills. To refer to this work as well as its result, Gruber coined a special term "conceptualization." Description that he called a "specification." Thus, ontology as proposed by Gruber defined as specification of a conceptualization.

Recently, an increasing number of theoretical and practical works has been published on multiattribute classification. The basic idea of current research in this area is the transformation of unstructured data sets in a

network of knowledge and the creation of tools and applications that work with them. The architecture of such systems is to be in the form of special knowledge bases (ontologies). The main feature of the ontological research is, in particular, the separation of the real world into components and object classes and determination of their ontology, or the aggregate of the fundamental properties that determine the changes in values and behavior. Principles of ontological research:

- Ontological analysis begins with a glossary of terms used in the discussion and study of the characteristics of objects and processes that make up the system in question, as well as a system of precise definitions of these terms. In addition, documenting basic logical relationships between the entered concepts. The result of this analysis is the ontology of the system, or set of vocabulary terms and precise definitions of the relationships between them;
- In any system, there are two general categories of objects of perception, such as the objects themselves, which constitute the system (physical and intellectual) and the relationship between these objects that characterize the state of the system. In terms of ontology, the concept of the relationship clearly describes (is an accurate descriptor of) the relationship between objects in the real world, and the terms are, respectively, the precise descriptors of actual objects themselves;
- Therefore, ontology is a data dictionary, which includes terminology and models the system behavior.

METHODOLOGY is the most popular and effective manual method to create ontology (Uschold, King, 1995). In this method Gruber's principles are realized and ontology specification software environment ODE (Ontology Design Environment) is developed. METHODOLOGY emphasizes the following procedures in the "life cycle" of ontology creation: the concepts vocabulary is created; it includes classes and its examples, attributes, activities, relationships etc. After that, concepts classification trees are created – main taxonomies are identified, where each taxonomy is a part of ontology being created (Ohgren, Sandkuhl, 2005).

Suggested approach

Step wise ontology engineering idea was proposed. It includes three steps:

- Vocabulary;
- Hierarchies;
- Relations.

The complete description of these steps is given in section of Implementation of suggested idea.

Tasks to be solved

There are two tasks to be solved:

- Flow-chart of the classification process construction as logical base for ontology construction (Rokach, Maimon, 2002);
- Construction of the domain independent classification method ontology.

Implementation of the suggested idea

Here is presented a case study of the method ontology construction. The first step is to create concept vocabulary. It is a set of concepts describing multiattribute classification method. Concept vocabulary includes classes of multiattribute classification method. Concept vocabulary can be seen in Table 1, there are 33 concepts included into the concept set.

Table 1

Concept vocabulary

Classes
Attribute, Splitting criterion, Class, Node label, Arc, Node, Stopping criterion, Classifier, Bias, Pruning, Attribute data type, Growth stopping criterion, Attribute type, Target attribute, Describing attribute, Discrete value, Continuous value, Process, Learning, Testing, Pruning, Root, Record, Internal node, Leaf node, Correctly classified records, Incorrectly classified records, Unclassified records, Growth of error, Data set, Learning set, Test set, Recognition accuracy.

The next step is to create taxonomies of these concepts. Formal form of the taxonomy is shown in Fig.1.

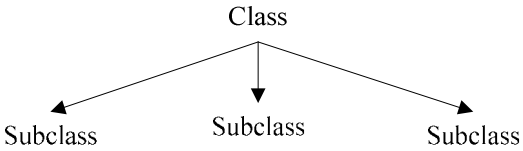


Fig.1. **Structure of the taxonomy.**

As a result, 18 taxonomies are created using classes of concept vocabulary, also new concepts were added (see Table 2). It allows detailing taxonomies structures and get a more accurate representation of concept hierarchies.

For example, concept "Node" was taken. It is the top class and it has the following subclasses – root, internal node, leaf node. "Node" concept taxonomy is built as shown in Fig. 2.

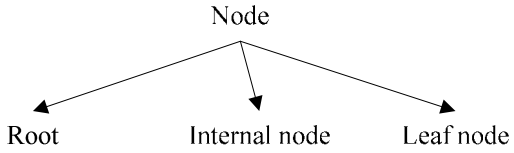
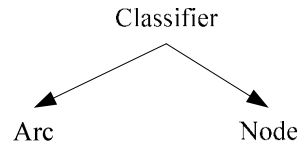
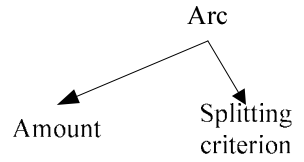
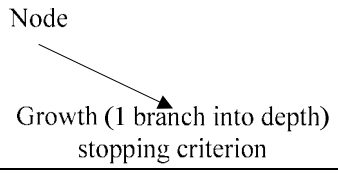
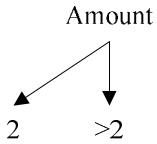
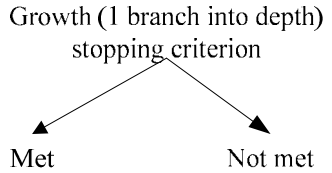
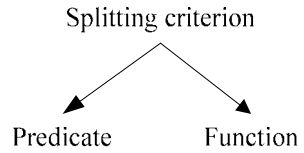
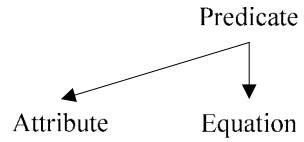
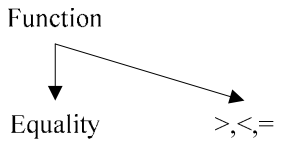
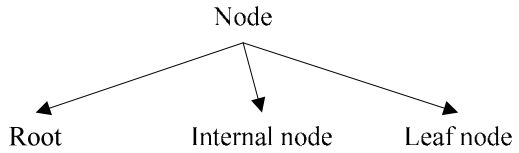
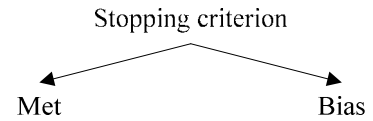
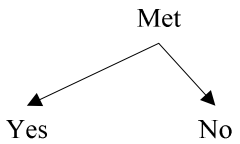
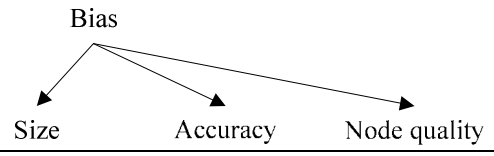
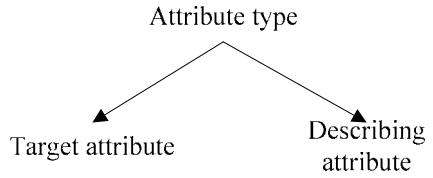
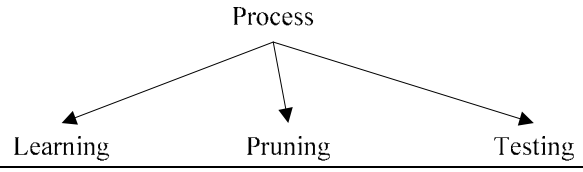
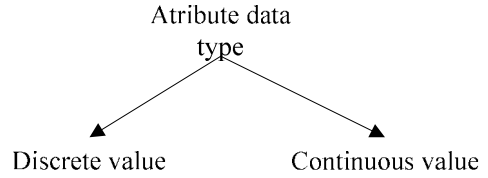
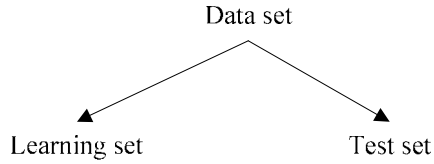
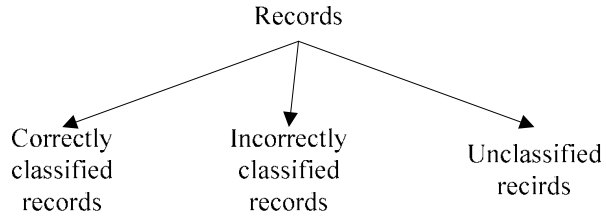
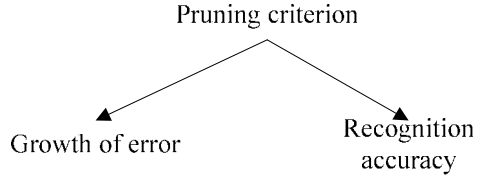


Fig. 2. **Taxonomy of the class "Node".**

Table 2

Concept taxonomies

Nr.	Used concepts	Taxonomy
1.	Classifier, Arc, Node	
2.	Arc, Amount, Splitting criterion	
3.	Node, Growth stopping criterion	
4.	Amount, 2, >2	
5.	Growth stopping criterion, Met, Not met	
6.	Splitting criterion, Predicate, Function	
7.	Predicate, Attribute, Equation	
8.	Function, Equality, "> < ="	
9.	Node, Root, Internal node, Leaf node	

10.	Stopping criterion, Met, Bias	
11.	Met, Yes, No	
12.	Bias, Size, Accuracy, Node quality	
13.	Attribute type, Target attribute, Describing attribute	
14.	Process, Learning, Pruning, Testing	
15.	Attribute data type, Discrete value, Continuous value	
16.	Data set, Learning set, Test set	
17.	Records, Correctly classified records, Incorrectly classified records, Unclassified records	
18.	Pruning criterion, Growth of error, Recognition accuracy	

The third step is to identify possible relationships between concepts. The formal form of these bonds is shown in the Fig. 3., where Concept 1 is the Object, but Concept 2 – the Subject.

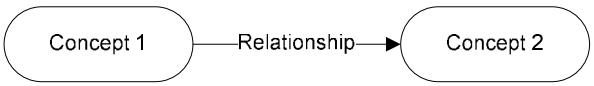


Fig. 3. **Relationship between two concepts.**

Relationships identifying step was divided into two substeps: creating of relationships set and relationships identifying between concepts. Relationships set includes: Is-a, Is part, Identify, Stops reaching, Characterize, Verify, Is result of, Use, Is compared with. After this step the possible relationships between any two concepts were identified. For example, we took relationship "Identify" and two concepts "Splitting criterion" and "Attribute". We know attribute at a tree node is selected using splitting criterion, respectively, bond "Splitting criterion – Identify – Attribute" is obtained. Remaining bonds were obtained similarly (Table 3).

Table 3

Relationships between concepts

Nr.	Relationships	Concept-Relationship-Concept
1.	Is a	Node – Is a – Attribute
2.	Identify	Splitting criterion – Identify – Attribute
3.	Stops reaching	Learning – Stops reaching – Stopping criterion
4.	Is result of	Rules set – Is result of – Learning
5.	Is a	Rule set – Is a – Classifier
6.	Verify	Testing – Verify – Classifier
7.	Characterize	Correctly classified records – Characterize – Classifier
8.	Characterize	Incorrectly classified records – Characterize – Classifier
9.	Characterize	Unclassified records – Characterizes - Classifier
10.	Is result of	Prognoses class – Is result of – Query
11.	Is compared with	Prognoses class – Is compared with – Real class
12.	Is part	Test set – Is part – Data set
13.	Is part	Learning set – Is part – Data set
14.	Use	Learning – Use – Learning set
15.	Use	Testing – Use – Test set

In the Fig. 4. binary relationships between concepts are seen.

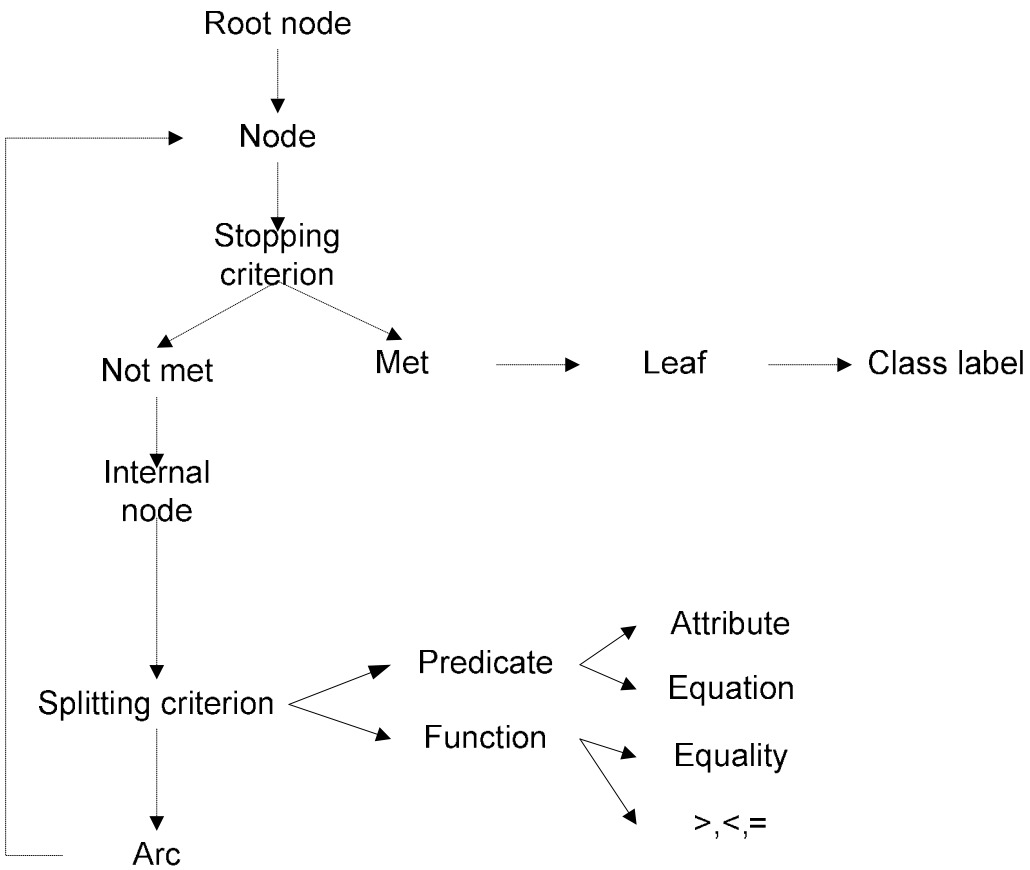


Fig. 4. **Binary relationships between concepts.**

In the end the method ontology was constructed, based on the information and data of the previous steps (Fig. 5).

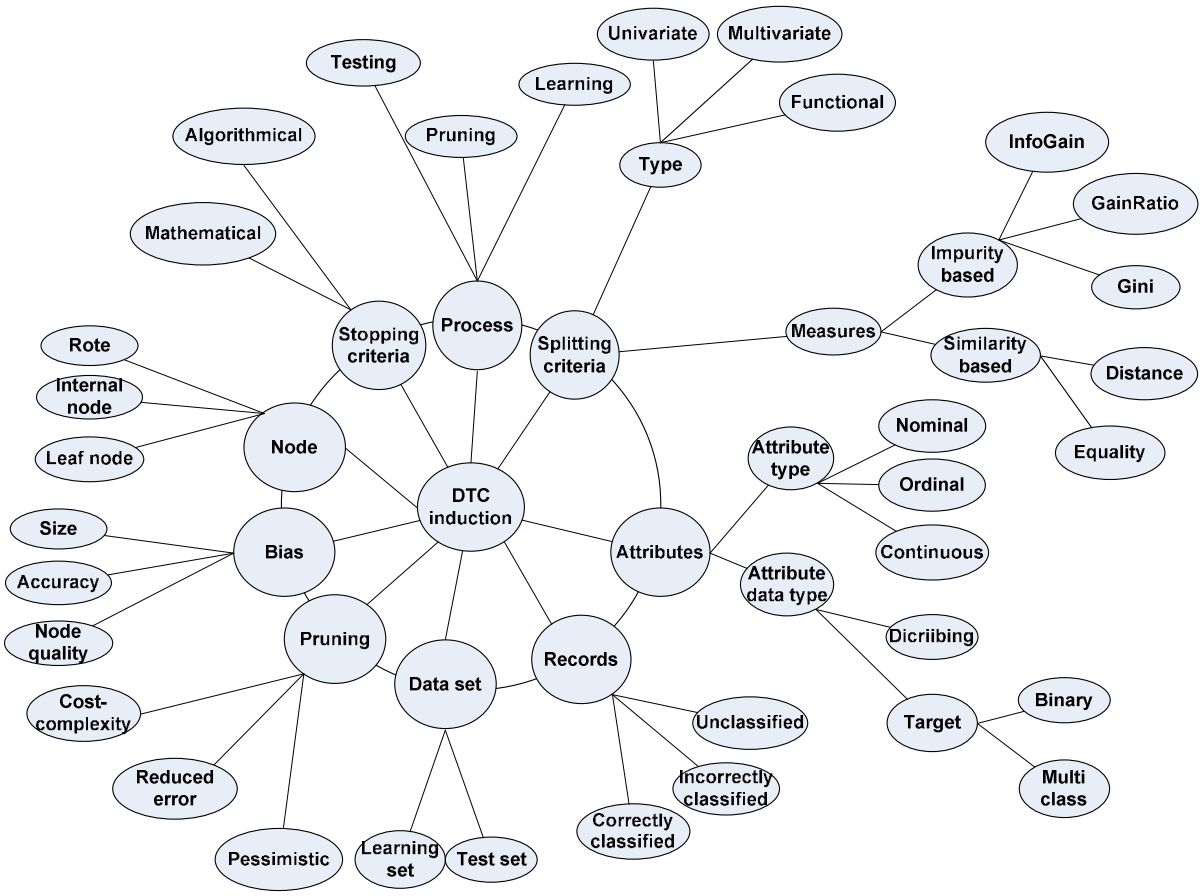


Fig. 5. **Method Ontology.**

Summary and future work

The investigation of domain ontology and method ontology interaction in the process of ontology-based classification system functioning. Problem Solving Method declare the format and semantics of the knowledge and data that they expect from the domain to perform their task. Problem Solving Method thus provides method ontology. As a future work XML, OWL and RDF will be studied. This will allow us to move towards the development of methods of applying ontologies.

References

Ben-David, D., Domany, T., Tarem, A., 2010. Enterprise Data Classification using Semantic Web Technologies. In: ISWC.

Gruber, T.R., 1991. The role of common ontology in achieving sharable, reusable knowledge bases. In: Allen J.A., Fikes R., Sandwell E. (eds) *Proceedings of the Second International Conference*, pp. 601-602.

Guarino, N., 1998. Formal ontology in information systems. In: Proceedings of FOIS'98, Trento, Italy, IOS Press, pp. 3-15.

Niaraki, A.S., Kim, K., 2009. Ontology based personalized route planning system using a multicriteria decisionmaking approach. In: Expert Systems with Applications, 36(2, Part 1), pp. 2250 – 2259.

Ohgren, A., Sandkuhl, K., 2005. Towards a Methodology for Ontology Development in Small andMedium-Sized Enterprises. In: *IADIS Conference on Applied Computing*.

Polaka, I., Tom, I., Borisov, A., 2010. Decision tree classifiers in bioinformatics. In: Scientific Proceedings of Riga Technical University, Information Technology and Management Science, Issue 5, Vol.44, RTU, Riga, P. 118-123.

Prabowo, R., Jackson, M., Burden, P., 2002. Ontology-Based Automatic Classification for the Web Pages: Design, Implementation and Evaluation. In: Proceedings of the 3rd International Conference on Web Information Systems Engineering (WISE.02).

Rokach, L., Maimon, O., 2002. Top-Down Induction of Decision Trees Classifiers. In: IEEE Transactions on systems, man and cybernetics: part C, vol. 1, 11.

Uschold, M., King, M., 1995. Towards a Methodology for Building Ontologies. In: Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95.

Optimization of automation decisions in regression testing

Anastasija Solovjova, Janis Grabis

Riga Technical University, Meza str. 1/4, Riga, LV-1048, Latvia
anastasija.solovjova@rtu.lv, grabis@rtu.lv

Abstract: *Regression testing is the testing of the previously tested program to ensure that changes did not cause the unexpected defects to the unchanged part of the software. Regression testing is the time-consuming process that also might be very expensive. That is why it is recommended to automate the test cases to archive the best result. By automating the test cases we can reduce the testing time and the testing cost, make the tests more accurate and more reliable, as well as give the possibility for testers to concentrate on the more difficult tasks then on the monotonous regression testing. However it is not possible and also not effective to automate all the test cases at once. The goal of the research is to create an optimization model that selects the test cases for automation. The model is based on the three main parameters – the testing time, the testing cost, the test case's priority. The goal of the optimization model is to reduce the testing time. Several experiments are done using one system as the example. The results received are compared with the random ones. As the result of the research there is the conclusion how effective is the optimization model introduced as well as the testing automation process is evaluated in general.*

Keywords: Regression testing, automation, optimization, test case selection.

Introduction

Software applications are frequently modified to meet changing user requirements. There is a risk of mistakes being made during the modification. As a result of the mistakes the software might not be working as expected. The unexpected behaviour often occurs in the unchanged part of the software. The regression testing is used to reduce such problems.

Regression testing is the testing of the previously tested program to ensure that changes did not cause unexpected defects to the unchanged part of the software (ISTQB, 2010). It is especially important in the iterative and incremental software development process because during such a development process the software is implemented piecemeal. Regression testing is also important during the software maintenance phase when it is necessary to introduce small changes or fix some defects. It is important not to break the other functionality while implementing such changes.

The regression testing is the object of many investigations (Agrawal et al., 1997; Biswas et al., 2011). Several methods are proposed to make the regression process more effective. The main research streams are regression test case creation and regression test case selection problems.

Because of the fact that regression testing is a time-consuming and repeatable process the good practice is to automate the regression testing. The automation of the regression testing reduces regressing testing time and cost, gives more accurate and faster results about the product quality as well as gives a possibility for human resources to concentrate on execution of more difficult tasks (Pettichord, 1999). It is clear that the automation of the test cases requires big investment for the test suite creation and maintenance. To ensure that the test cases automation is successful, it is important to define a subset of the test suite for automation.

The goal of this paper is to elaborate an optimization model that selects the test cases for automation. The optimization objective is to reduce the testing time. The optimization model should meet the following requirements:

- to observe the cost restrictions;
- to provide the sufficient coverage level;
- to observe the test cases priorities that are determined by the test cases risk levels.

The rest of the paper is organized as follows. The first section provides the view of the related literature. The second section describes the proposed optimization model. In the fourth section the experiments description and result is presented. In the fifth section the conclusion is given.

Literature review

Regression testing and efficiency of regression test suite is an active research area, and many methods have been proposed. All techniques can be grouped as follows:

- test cases prioritisation techniques;
- test cases selection techniques;
- test cases reduction techniques.

The test cases prioritisation techniques leave the full test suite, but change the order of the test cases to find the defects as soon as possible. The test cases selection techniques select the subset of the test suite according to specific criteria. For example, the subset of the test cases that tests only modules that were affected by changes. The test cases reduction techniques reduce the test suite by combining the test cases that test the same modules or by determining and removing the redundant test cases. The test cases selection techniques are similar to test cases reduction techniques. The main difference is that the goal of the test case reduction techniques is to provide the best coverage for the specific software version by the minimal number of the test cases. While the test cases selection techniques focuses on the changes between the previous and current software versions.
Harman and Yoo (2007b) review four test cases prioritisation techniques:

- coverage-based;
- distribution-based;
- human-based;
- requirement-based.

The coverage-based prioritization techniques first-and-foremost apply test cases that provide the best coverage. The distribution-bases techniques divide the test cases into clusters. That helps to identify the test cases that do similar or even the same things. The human-based techniques determine the test cases priority based on human decisions. The requirement-based techniques set the priority of the test cases according use cases, client requirements and the complexity of the implementation.
Researchers (Harman et al., 2007a) mention the greedy algorithm as one of the possible methods to prioritise the test cases. The authors (Harman et al., 2007a) also propose an advanced greedy algorithm. This algorithm works similar like the traditional greedy algorithm, but remembers the previous choice. For example, the algorithm at first selects the test case which covers the most statements. On the next step it finds test case with the next greater weight and compares the test case with the already chosen one. The test case that tests the largest number of the statements that are not covered with the test cases yet is selected next.
One more algorithm that is reviewed by Harman et al. (2007b) for test cases prioritization is the hill climbing algorithm. While applying this method to regression testing it is considered that there are n-1 neighbours for each test case, if n is the number of the test cases.
Gu et al. (2010) propose to use the heuristic algorithm not to prioritize the test cases but to reduce the test suite size. This algorithm has the following steps:

1. select the test cases from the test suite that covers the same requirements;
2. select the only one test case from the obtained set that tests specific requirement;
3. in case if several test cases test the same requirement then select the only one using the greedy algorithm;
4. repeat the previous steps for each requirement considering the coverage that are archived by the previously selected test cases.

Kumar et al. (2011) propose to reduce the test suite size by combining the test cases that verify the functionality with the test cases that verify boundary values. The algorithm combines the test case reduction and prioritization techniques. The proposed steps are as follows:

1. identify if the functionality and the boundary values might be tested together;
2. identify the situations when the functionality and the boundary values might be tested using the single test case;
3. prove that single test case covers both aspects;
4. select the test cases from the obtained test set that cover the happy path and set them the highest priority.

Graves et al. (2001) review the regression test case selection techniques. The most techniques might be grouped as follows:

- dataflow techniques;
- save techniques.

The dataflow techniques are based on the dataflow coverage and select the test cases that test the modified data interaction. These techniques require that each definition-use pair is tested. The safe techniques guarantee that set of the test suite contains all test cases form the original test suite that might find the defects in the modified software.
Although previous studies acknowledge importance of regression testing automation, there are few formal guidelines for selecting test cases for automation. Having such guidelines is important to balance investments required for test case automation and testing time savings due to automation.

Model

The aim of the proposed optimization model is to determine the test cases from the manual test suite that should be automated to reduce the testing time. The optimization model is developed according to best practices of regression testing and accumulated empirical experiences. The objective function of the model is as follows:

$$Z = \sum_{\forall i}(\tau_{m_i} \times x_i + \tau_{a_i} \times y_i + (R_i - 1) \times \frac{1}{2} \times \tau_{m_i} \times x_i) \rightarrow min, \quad (1)$$

where τ_{m_i} – manual testing time for the test case i, time units;
τ_{a_i} – automated testing time for the test case i, time units;
x_i – variable that shows if it the test case i should be left manual, $x \in (0,1)$;
y_i – variable that shows if it the test case i should be automated, $y \in (0,1)$;
R_i – the priority of the test case i.

The test cases are selected according to their priority. The priority of the test case is determined by the test case risk level. The model favours automation of test cases with higher priority. That is represented by using the penalty function $(R_i - 1) \times \frac{1}{2} \times \tau_{m_i}$. The coefficient $R_i - 1$ is set because for test cases with the lowest priority $R_i = 1$ no penalty function is applied. The penalty time is applied only for the manual testing time to maximize the automation probability for the most risky test cases. The coefficient $\frac{1}{2}$ is used because if the manual test execution does not discover the existing defect the testing should be repeated and at least half of the provided time should be spent.

The regression testing is done regularly. The regularity of the regression test cases execution depends on the number of the testing cycles n.

Each test case i tests one or several procedures j. The traceability between the test cases and procedures is defined using the binary attribute K_{ij}, $K \in (0,1)$.

Four constrains are defined in the model. The constraint (2) says that each procedure should be tested by at least one test case. The constraint (3) limits the cost of the automation. The constraint (4) says that the test case might not be automated and left manual at the same time. The constraint (5) excludes the test cases that do not test any procedure.

$$\sum_{\forall i,j}(x_i + y_i) \times K_{ij} \geq 1 \; \forall j \quad (2)$$

$$Ca(n) + Cm(n) \leq \gamma \times Conlym(n), \quad (3)$$

where $Ca(n)$ – the testing cost for automated test cases after the optimization, monetary units;
$Cm(n)$ – the testing cost for the manual test cases after the optimization, monetary units;
$Conlym(n)$ – the testing cost before the optimization, monetary units;
γ – the coefficient that says how many times the $Ca(n) + Cm(n)$ might be greater then $Conlym(n)$.

$$x_i + y_i \leq 1 \; \forall i, \quad (4)$$

$$x_i + y_i = 0 \; \forall i, if \sum_{\forall i} K_{ij} = 0, \quad (5)$$

The testing time consists of:

- test case creation time;
- test case execution time;
- test case execution result's analysis time;
- test case maintenance time.

The manual testing time for each test case is calculated using the equation (6).

$$\tau_M(n) = \tau_{M_{ex}} \times n + \tau_{M_{cr}} \times p \times (n - 1), \quad (6)$$

where $\tau_{M_{ex}}$ – manual execution time, time units;
$\tau_{M_{cr}}$ – manual test case creation time, time units;
p – percentage that shows test case maintenance time subject to the test case creation time.

It is assumed that the manual test suite is already ready. Therefore for manual test cases, the test case creation time is equal to 0. However to calculate the test case maintenance time it is important to know how much time was spent on the test suite creation. The maintenance time is calculated as the percentage of the test case creation time. On the first testing cycle the test case does not require the maintenance. That is why the maintenance time for the testing cycle n is multiplied by the $(n - 1)$. For the manual test cases it is assumed that the result analysis time is not required.

The automated testing time for each test case is calculated using the equation (7).

$$\tau_A(n) = \tau_{A_{cr}} + \tau_{A_{cr}} \times p \times (n - 1) + (\tau_{A_{ex}} + 0,5 \times \tau_{A_{an}}) \times n, \quad (7)$$

where $\tau_{A_{ex}}$ – execution time of the automated test case, time units;
$\tau_{A_{cr}}$ – automated test case creation time, time units;

Test case execution result's analysis time depends on the test case execution status. If the test is passed then no time is needed for analysis. If the test is failed then the time is required to analyse the reason of the failure. It is assumed that the test might fail and pass with the same probability equal with 0,5.
While calculating the testing time for the automated test cases, the test case relationship is taken into the account. The test case relationship means that one test case uses the same functions as another test case. If the test case T1 is automated and the test case T2 might reuse the functions already automated in the test case T2 then the automation time of the test case T2 might be less than in case if T1 is not automated. The time of the creation the automated test case is calculated using the equation (8).

$$\tau_{A_{cr_i}} = Min(\tau_{A_{cr_1}} - 0{,}8 \times \tau_{A_{cr_q}} \times U(i,q) \times y_i)\ \forall i, \forall q, \tag{8}$$

where $U(i,q)$ – binary test case relationship's attribute that shows if test case i might be based on the test case q, $U \in (0,1)$.
In case if test case i might use for basis many test cases then the test case with the maximum creation time is taken as the test case q. The coefficient 0,8 is used because there might be necessary to adapt test case for using another test case.
The total testing time is calculated using the (9) equation. The manual time spent for testing after the automation is introduced is calculated using the (10) equation.

$$\tau_{total} = \sum_{\forall i} \tau_{m_i} \times x_i + \tau_{a_i} \times y_i \tag{9}$$

$$\tau_{mTotal} = \sum_i \tau_{M_i}(n) \times x_i + \sum_i (\tau_{A_i}(n) - \tau_{A_{i_{ex}}} \times n) \times y_i, \tag{10}$$

The equation (11) is used to calculate the testing cost before applying the optimization.

$$Contym = S_e \times \sum_i \tau_{M_i}, \tag{11}$$

where S_e – average salary per person hour, monetary units;
Equation (12) is used for calculation the cost of manual testing after the optimization. Equation (13) shows the cost of the automated testing.

$$Cm = S_e \times \sum_i \tau_{M_i}(n) \times x_i, \tag{12}$$

$$Ca = S_e \times (\sum_i (\tau_{A_i}(n) - \tau_{A_{ex_i}} \times n) \times y_i + \tau_{A_{conf}}) + Cr + Ct, \tag{13}$$

where Ct – cost of the automation tool licence or/and modification, monetary units;
Cr – cost of the human training, monetary units;
$\tau_{A_{conf}}$ – time required for the tool configuration, time units.
The attribute $\tau_{A_{conf}}$ is not multiplied by the number of the testing cycles n because it is assumed that tool configuration is the one-time activity.

Experimental

The model elaborated was evaluated using a practical example and characteristics of regression testing automation are analysed experimentally.

Case description

The system used in experiments is a web application used in the financial sector. The application consists of 25 screens and has approximately 1135 data input fields. The size of the application and complex business logics makes the manual regression testing infeasible. The testing is partly automated, i.e. the data input is tested for some screens. However no specific business rules are verified by the automated tests. The testing that is used in the project is the black box testing. The test case describes the steps, input values and the expected result. One test case tests one or several procedures. The procedure is a use case or a specific business rule. One functional part of the system is used to evaluate the model. The number of the test cases used in the experimental studies is 16, and the number of the procedures to be tested is 21.

Experimental design

The input data for the optimization model are prepared in multiple steps. At the first step it is needed to create the traceability matrix between the test cases and the procedures that they test. The second step is to determine the priority of each test case. The priority of the test case is set by the testers based on the most important business scenarios. The test case priority is set in scale from 1 to 4 where priority "4" is the most important. On the third step the test cases relationship matrix was created.
The input values for the testing time calculation were prepared during the next step. The process was as follows:

- the manual test cases were executed and the time for each test case execution was recorded;
- time that was spend for the test cases creation was divided by the number of the test cases;
- the test case point analysis method (Govindrajan et al., 2001) was used to forecast the time required for the automation of the test cases;
- the existing automated test cases were analyzed to determine the time for execution of each action.

Based on the result, execution time of the automated test cases was forecasted.
The following experiments are planned:

- to check if the optimization model gives better result than the existing process;
- to check if the optimization result is better than the random generation model;
- to evaluate influence of the number of testing cycles on automation decisions;
- to evaluate influence of the cost coefficient γ on automation decisions.;
- to evaluate influence of person hourly rates on automation decisions.

Results

The existing process parameters were calculated using the summarized data. The data was compared with the optimization result. The data is summarized in the Table 1.

Table 1

Experiment result

Parameter	**Before optimization**	**Model execution result**
Objective Z	151,93	81,98
Total testing time τ_{total}	75,95	81,98
Manual testing time τ_{mTotal}	75,95	56,86
Number of the manual test cases	16	1
Number of the automated test cases	0	14
Total number of the test cases	16	15
Highest priority test cases automated	0	5
Total testing cost $Ca(n) + Cm(n)$	265,83	286,52

Firstly, we can see that the number of test cases used is reduced by one test case. This means that other test cases cover the same procedures as the redundant test case. Secondly, all the highest priority test cases are automated. That reduces the risk of not finding all errors. Thirdly, the objective's value is reduced almost 2 times.
The negative aspect of results is that the total testing time and total testing cost have increased by 8 percent. However, the manual testing time is reduced by the 25 per cent, and the tests are still completed within the allocated time interval.
The optimization results were compared with random one. The random selection model chooses test cases to be automated by chance. Based on the result the objective and other values were calculated for the random choice. The results in Table 2 are reported relative to the results given by the optimization model.

Table 2

Relative increase of testing performance measures in the case of random selection

Experiment	Z	Number of the manual test cases	Number of the automated test cases	Number of the automated 4th priority's test cases	τ_{total}	τ_{mTotal}	$Ca(n) + Cm(n)$
1	1,85	16	0	0	0,93	1,34	1,14
2	1,11	0	1,14	1	1,11	1,08	1,05
3	1,61	9	0,50	0,4	1,03	1,34	1,19
4	1,39	5	0,79	1	1,17	1,37	1,23
5	1,20	2	1	0,8	1,13	1,18	1,11
6	1,78	13	0,21	0	1,06	1,43	1,23
7	1,81	11	0,36	0,2	1,13	1,42	1,23
8	1,88	15	0,07	0	1,02	1,43	1,22
9	1,08	1	1,07	1	1,08	1,09	1,06
10	1,66	7	0,64	0	1,12	1,27	1,15
Average Values	1,54	7,9	0,58	0,48	1,08	1,29	1,16

The testing time and the cost obtained with the random selection model are greater than those given by the optimization model. The testing suite is not reduced and not all the highest priority test cases are automated.

To evaluate impact of the number of testing cycles on automation, 20 experiments were performed by varying $n = 1$ till $n = 20$. The graphical interpretation of the result is shown on the Fig. 1. The overall tendency is that the level of automation increases with the increasing n. On the 9th testing cycle we can see that the automation level was reduced. This exception is explained by the test cases relationship. The test case is automated not only because automation improves particular test case's testing time, but also because it automation can reduce other test case testing time.

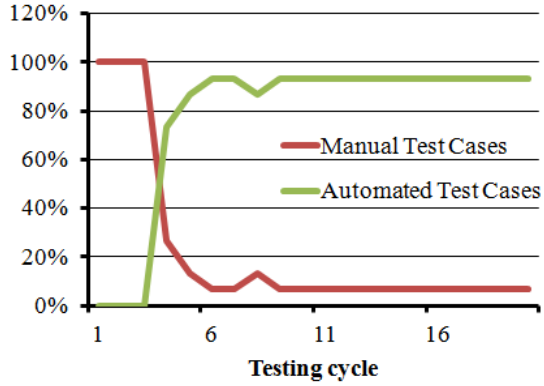


Fig. 1. **Automation level dependency on the number of testing cycles.**

The testing cost dependency on the number of testing cycles is shown on the Fig. 2. The larger is number of the testing cycles, the quicker the automation pays back. Also we see that the automation gives the effect only in case if there is sufficient number of the test cases.

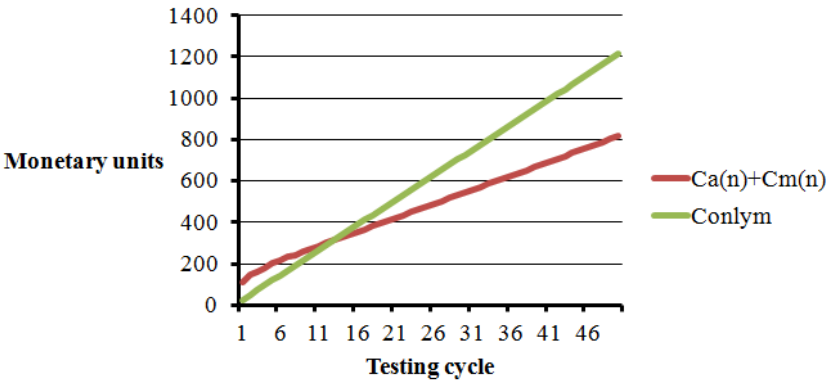


Fig. 2. **Testing cost dependency on the number of testing cycles.**

To evaluate influence of the cost coefficient on the automation, several experiments were performed by varying $\gamma = 1$ till $\gamma = 5$ with the interval = 0,5 and $n = 1$ till $n = 10$. The results are summarized in the Table 3. The cost coefficient influences the automation level only on the first testing cycles. This means that it is possible to optimize the process even if there are a few testing cycles, but this requires larger investments in automation.

Table 3

The automation level subject to number of the testing cycles and cost coefficient

n	γ				
	1	2	3	4	5
1	0	0,00	0,00	0,00	0,00
2	0	0,00	0,33	0,53	0,53
3	0	0,00	0,47	0,47	0,47
4	0	0,73	0,73	0,73	0,73
5	0	0,87	0,87	0,87	0,87

To evaluate if there is a dependency between the automation level and the rate per person hour, the experiments with the different value of the S_g parameter were performed. The graphical interpretation of the results is shown in the Fig. 3. If the rate per person hour is high then it is better to use the greater level of the automation.

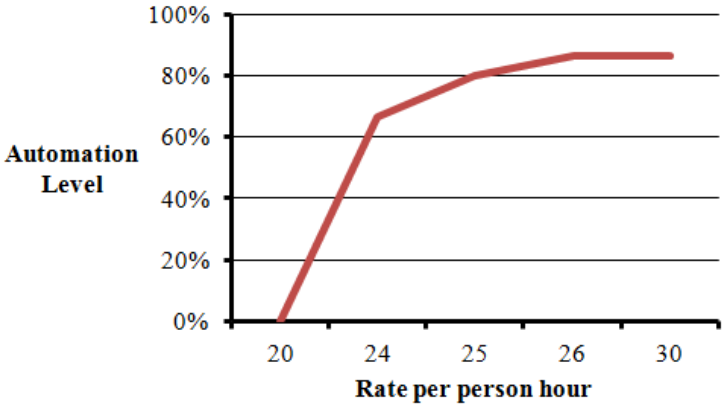


Fig. 3. **Automation level dependency on the rate per person hour.**

Conclusion

In this article the optimization model for the test case selection was presented. The model:

1) shows which test cases from the manual test suite is better to automate;
2) reduces the test suite size;
3) considers the relationships among the test cases;
4) gives possibility to limit automation costs.

The model was evaluated using the practical example, and it is showed to outperform both the random test case selection approach and the currently existing automation approach at the company.

Several experiments were performed to identify opportunities for improving the test case automation. Based on the results of the experimental analysis, the following conclusions are drawn:

1) more test cases should be automated as the number of testing cycles increases;
2) the larger rate per person hour requires a higher level of automation;
3) while automating the test cases it is worth not only to check if particular test case automation improves the result, but also to evaluate how the automation of the particular test case influences other test cases;
4) if the number of testing cycles is large it is also possible to optimize the testing process using automation, but this requires significant financial investment;
5) if the number of testing cycles is small then automation is not effective.

References

Agrawal, H., Horgan, J., London, S., Wong, W. A., 1997. Study of Effective Regression Testing in Practice // 8th IEEE International Symposium on Software Reliability Engineering (ISSRE'97), Albuquerque, NM, November 1997, pp. 264-274,

Biswas, S., Mall, R., Satpathy, S., Sukumaran, S., 2011. Regression Test Selection Techniques: A Survey // Informatica 35, 2011, pp. 289–321.

Graves, T., Harrold, M., Kim, J., Porter A., 2001.An Empirical Study of Regression Test Selection Techniques // ACM Transactions on Software Engineering and Methodology – Vol. 10, No. 2, April 2001, pp. 184–208.

Govindrajan, M., Maharana, S., Patel, N., Ramdas, S., 2001. Test Case Point Analysis // White Paper Version 1.0, April 11 2001, 13 p.

Gu, Q., Tang, B., Chen, D., 2010. Optimal Regression Testing based on Selective Coverage of Test Requirements // International Symposium on Parallel and Distributed Processing with Applications (ISPA 10), September 2010, pp. 419-426.

Harman, M., Hierons, R., Li, Z., 2007. Search Algorithms for Regression Test Case Prioritization // IEEE Transactions on Software Enfineering, Vol. 33, No. 4, April 2007a, pp. 225-237.

Harman, M. and Yoo, S., 2007. Regression Testing Minimisation, Selection and Prioritisation : A Survey // Software Testing, Verification, Reliability, Published online in Wiley InterScience (www.interscience.wiley.com), 2007b, pp. 1–7

ISTQB – Veenendaal, E. van (ed.) Standard glossary of terms used in software testing v2.1. // International Software Testing Qualifications Board, dd. April 1 2010, 51 p.

Kumar, K., Pradesh, A., Rao, A., 2011. An Approach to Cost Effective Regression Testing in Black-Box Testing Environment // International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1 (May 2011), pp. 198-208.

Pettichord, B.,1999. Seven Steps to Test Automation Success // STAR West, San Jose, November 1999, 17 p.

Quantitative modelling of qualitative knowledge in the evaluation of stakeholder influence

Ojaras Purvinis, Dalia Susniene

Kaunas University of Technology, Klaipedos 3, KTU, Panevežys, Lithuania
opurvi@panf.ktu.lt, dalia.susniene@ktu.lt

Abstract: *Usually verbal facts or attitudes drawn from questionnaires are approximate and contain uncertainty from the mathematical viewpoint. The paper demonstrates how the fuzzy logic could be applied do develop a quantitative model from qualitative knowledge in business stakeholder case.*

Keywords: fuzzy logic, computer simulation, stakeholder influence.

Introduction

There exist a lot of mathematical, statistical methods and computer modelling techniques for the researcher when the numerical data of the object are available.
Modelling and computer simulation can contribute to a deeper understanding of the issue (Scheffran, 2006). To develop a model that enables to explain or predict what could happen if some parameters are changed usually one needs precise mathematical and statistical estimation of these parameters, input and output variables.
Economics and other social sciences usually employ regression and correlation analysis for investigation of the connections between numerical factors (Fig.1). The scatter diagram visualizes the statistical dependence between factors, the correlations coefficient *R* numerically evaluates the strength of dependence and regression equation can be used for the prediction of the value of factor *Y*.

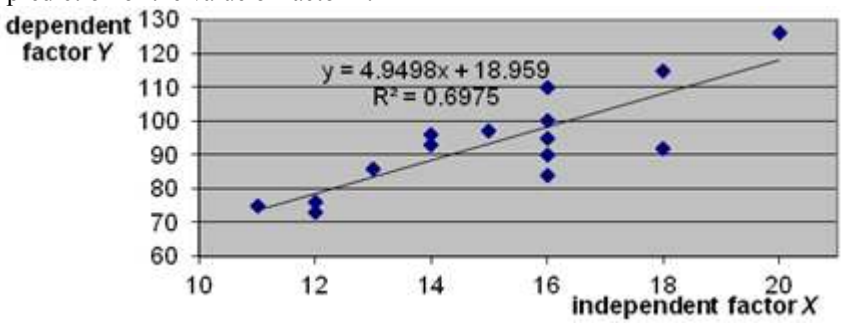


Fig. 1. **Scatter diagram, correlation coefficient and regression line**.

But it is not the case when the knowledge and verbal conclusions of experts or answers of questionnaire respondents about the influence of input factors or parameters on output are not numerical (Purvinis and Šukys, 2012). Usually such verbal facts or attitudes are approximate and contain uncertainty from the mathematical viewpoint.

The application of fuzzy logic

For instance, we need to evaluate the reliability of a possible business stakeholder. The most cited stakeholder definition was proposed by Freeman (1984) which included those groups or individuals who are affected by the organization as well as those who can affect it (Freeman and McVea, 2001, 16; Bryson et al., 2002). Stakeholders are an integral part of the business environment and this part should be managed by evaluating their influence. It cannot be measured and expressed digitally. The qualitative verbal evaluation of the influence can be low, medium or high. The same apply to other social sciences. For instance, about the possibility of the graduate student to find a job within, say one year, also one can say that it can be low, medium or high.
To overcome differences between qualitative and precise quantitative models and to build a digital model based on qualitative estimation of independent factors on the specific dependent factor authors of this paper suggest to apply fuzzy logic. Fuzzy logic is a precise logic of imprecision and approximate reasoning (Zadeh, 2008).
The fuzzy sets and fuzzy logic were introduced by prof. Lofti A. Zadeh in his research papers (Zadeh, 1965 and 1989). Over the years since these ideas were published, his proposal has gained recognition in various fields of the research (Celikyilmaz, 2009).
This paper demonstrates the idea to apply fuzzy logic method to develop a model of the influence of some hypothetic business stakeholder on the another enterprise stakeholder. Let's take into account such attributes of the business stakeholder as interest, power and reliability and evaluate their influence on the enterprise (Purvinis and Susnienė, 2013).

In order to give quantitative evaluation of these factors, each factor was divided into three linguistic levels, called terms in fuzzy logic,– *interest, reliability* and *power* into levels low, medium and high, and *influence* into negative, zero and positive. These levels enabled to code expert-type knowledge into IF – THEN rules, for instance:

IF interest is medium
and power is high
and reliability is low
THEN influence is medium.

It is easy to acquire such rules from questionnaires (Table 1).

Table 1

The questionnaire

***Interest* is**	***Power* is**	***Reliability* is**	**Please evaluate the *influence* in terms *low*, *medium* or *high***
low	low	low	?
low	low	medium	?
…	…	…	…
high	high	high	?

Taking into account that the IF part, called antecedent of the rules, was comprised of three factors interest, power and reliability where each had three levels, then the rule base consisted of $3^3 = 27$ rules.

The problem of knowledge uncertainty can be conveyed and discussed in details following the case of factor *reliability*. The factor was split into three linguistic levels – *low, medium* and *high* and besides that, the factor was assigned scores x ranging from 0 (lowest) to 10 (highest). This implied that scores were also divided into three subsets corresponding to above mentioned levels – *low, medium* and *high* (Fig. 2). In traditional set theory, a clear distinction exists between members and non-members of a set. Contrary to crisp sets, a fuzzy set or subsets, as in our case, can partially overlay each other and thus can be formed by assigning a membership value $m_{level}(x)$ to each level in the interval of [0,1] (Celikyilmaz, 2009). If, for instance, reliability was evaluated with score $x = 7$ and was calculated $m_{high}(7) = 0.6$, then this means that such interest belongs to the level *high* with membership value 0.6. It should be understood that the assertion „x is high“ is true just with degree 0.6, while the absolute truth has degree 1. The statement that is absolute false has the degree of truth equal 0, for instance, the statement „$x = 7$ is low“ is absolutely false, hence $m_{low}(7) = 0$.

Fig. 2 shows that membership functions overlap. For instance, when the score x=4, then such *reliability* is medium with the truth $m_{medium}(4) = 0.7$ and is low with the truth $m_{low}(4) = 0.3$ (Fig. 2). Hence, if the evaluation x of the factor *reliability* slightly changes or two experts give close to 4 but different scores, then anyway the reliability evaluation as partially low ant partially medium remains right, just slightly change the degree of truth of both evaluations. This enables to overcome the problem of imprecision of qualitative knowledge.

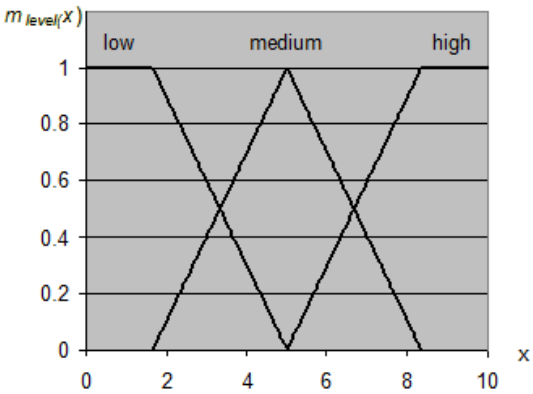


Fig. 2. **Membership functions $m_{low}(x)$, $m_{medium}(x)$ and $m_{high}(x)$ of the factor *reliability*.**

Similarly for each factors *power*, *interest* and the output factor *influence* membership functions $m_{low}(x)$, $m_{medium}(x)$ and $m_{high}(x)$ were also constructed. Neighbouring membership functions were overlapped. This assured compatibility of imprecise qualitative knowledge.

Thus the general structure of the fuzzy inference system took the form as shown in Fig. 3. There the input factors are shown on the left side. Values of these factors are processed by the rules block and performed by fuzzy interface. And finally, the digital value of the *influence* is obtained by the defuzzification

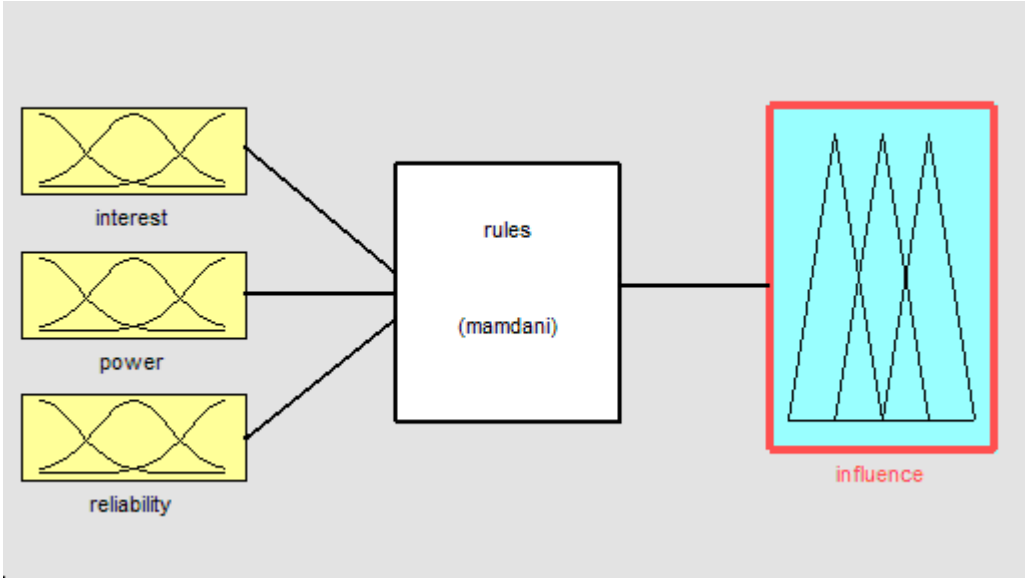


Fig. 3. **The fuzzy system**.

The relationship between input and output is called inference system. The inference is based on IF-THEN rules but it is a sophisticated process as the each input factor may simultaneously partially belong to two fuzzy levels. This implies that three input factors may simultaneously activate up to 2^3 rules with different contribution to the output factor *influence*.

Several methods exist to compute the output variable values (Celikyilmaz, 2009). This research was based on the convenient Mamdani inference method as it suited well for the types of membership functions used for factor *influence* fuzzy levels.

The developed model enables to explore the connection between qualitative input and output factors and to visualize this dependence in 3D space (Fig. 4). For instance, in given figure there is plotted an *influence* dependence on *interest* and *power,* while the factor reliability is fixed at the level *reliability* = 4.5 (partially *medium*, partially *low*). This demonstrates that modelling by fuzzy logic, factors can be assigned concrete scores or verbal levels as well. It is seen that at this fixed level of *reliability* the maximal influence is about 6 score points and it is reached when *interest* is high about 9-10 scores and does not depend on *power*. But it is the case when *power* is high. For all *power* values the *influence* remains dependent on *interest* and the *influence* increases along with *interest* increase. The plot also reveals that interest and power interact resulting in maximal values of the *influence*.

Similarly the model enables to present and to analyse the dependence of the *influence* on any two factors when the remaining third factor is fixed at the arbitrary level.

Discussion and conclusions

It should be pointed out that expert rules were just verbal and at no stage of the model development there were no explicit quantitative dependence of the output factor *influence* on the formulated input factors *interest*, *power* and *reliability*. Nevertheless the model, as it is seen in the Fig. 4, is quantitative and can be used in more general mathematical models.

Many fuzzy logic modelling environments, for instance such as Matlab Fuzzy add-on or FuzzyTech, have a capability to export model into programming language stand-alone code. This code can be later used in another more general model. For instance, in agent-based models the fuzzy submodel code may be used by intellectual agents, say suppliers or firms, in decision making or for evaluation of the dynamic business environment.

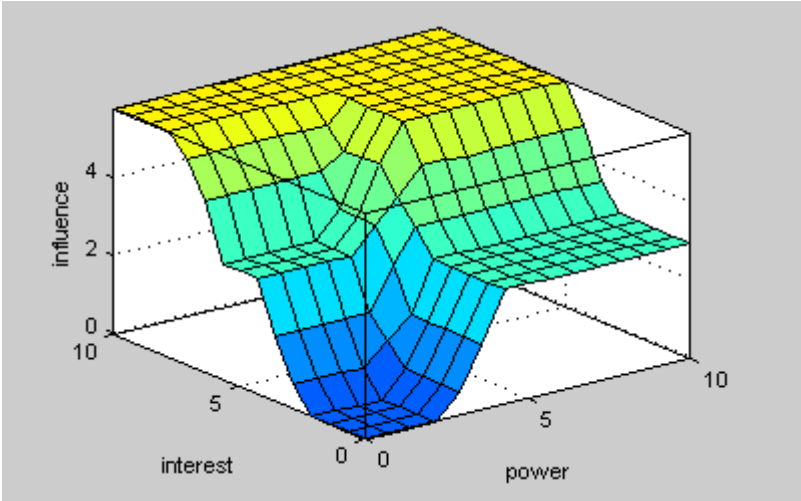


Fig. 4. **Plot of the *influence* dependence on *interest* and *power* when the factor reliability is fixed: on the level *reliability*=4.5 (partially *neutral*, partially *low*).**

The analysis of dependence plot also enables to find possible errors or contradictions in expert knowledge given in the rule form. If two input parameters are fixed and one is left to change, say *interest,* then the corresponding dependence *influence* (*interest*) should form a monotonous (increasing in this case) curve on the surface. If appears that the curve is not monotonous, i.e. the surface has a local minimum or maximum, then there should be an inconsistence in the rule base.

References

Bryson, J.M., Cunningham, G.L., Lokkesmoe, K.J., 2002. What to do when stakeholders matter: the case of problem formulation for the African American men project of Hennepin county, Minnesota, *Public Administration Review* 62, pp.568-584.

Celikyilmaz, A., Türksen, I.B., 2009. *Modeling uncertainty with fuzzy logic.* Springer-Verlag - Berlin Heidelberg, 400 p.

Freeman, R. E., 1984. *Strategic Management: A Stakeholder Approach.* Boston - Pitman Publishing, 276 p.

Freeman, R.E., McVea, J., 2001. A stakeholder approach to strategic management. In: M. Hitt, E. Freeman, and J. Harrison (eds.), *Handbook of Strategic Management*. Oxford - Blackwell Publishing, pp. 189-207.

Kleemann,Martin Welp, Stakeholder Dialogues. In: *Natural Resources Management: Theory And Practice*. Berlin - Springer, pp. 153-186.

Purvinis, O. and Susniene, D., 2013.*Understanding Stakeholder Influence for Successful Stakeholder Management.* (to appear).

Purvinis, O. and Šukys P., 2012. *Neraiškioji logika ir jos praktiniai taiky*mai. Kaunas - Technologija, 56 p. (In Lithuanian)

Scheffran and Jürgen., 2006. Tools for Stakeholder Assessment and Interaction. In: S. Stoll-Kleemann, M. Welp(Eds.), *Stakeholder Dialogues in Natural Resources Management: Theory and Practic*e, Springer, pp.153-185.

Zadeh, L.A., 1965. Fuzzy sets**.** *Information and control*, 8, pp. 338-353.

Zadeh, L.A., 1989. Knowledge representation in fuzzy logic. *IEEE Transactions on Knowledge and Data Engineering* 1, pp. 89–100.

Zadeh, L.A., 2008. Is there a need for fuzzy logic? *Information sciences*, 178 (13), pp. 2751-2779.

Perspectives of 1,4-DHP-lipid Molecular Dynamics

Vita Duka[1,2], Cezary Czaplewski[3], Adam Liwo[3], Gunars Duburs[2], Inta Liepina[2], Anda Zeidmane[1]

[1] *Latvia University of Agriculture, Liela str. 2, Jelgava, LV3001, Latvia*
[2]*Latvian Institute of Organic Synthesis, Aizkraukles str. 21, Riga LV1006, Latvia*
[3]*University of Gdansk,ul. Sobieskiego 18, 80-952 Gdansk, Poland*
vita.duka@llu.lv ; inta@osi.lv; anda.zeidmane@llu.lv

Abstract: *This paper focuses on software aided methodology of molecular dynamics using classical modeling cycle that can be used for investigation of complex lipid system molecular dynamics such as the cationic amphiphilic lipid type compound 1,1'-{[3,5-bis(dodecyloxycarbonyl)-4-phenyl-1,4-dihydropyridin-2,6-diyl]dimethylene} bispyridinium dibromide (1,4-DHP lipid). We summarized and systematized the molecular dynamics simulation process, and provided a list of software tools that can be successfully used for different purposes in different steps of molecular dynamics simulation. We show that molecular dynamics simulation as a computer modeling method complies with the assumptions of the mathematical modeling cycle. That was proved with the successful molecular dynamics studies of 1,4-DHP lipid system. The cycle of 1,4-DHP lipid system modeling was accomplished with the verification of results, that is excellent result although it opens a perspectives for further analysis of this system.*

Keywords: molecular modeling, molecular dynamics, lipid, gene transfection agent.

Introduction

Molecular modeling, also called computational chemistry, is the science that studies molecular structures through model building and uses principles of computer science to assist in solving chemical problems. Molecular modeling encompasses variety of computer based methods in order to understand and predict the behavior of molecular systems at the molecular level. This is a rapidly progressing area due to the development of theoretical methods based on classical, statistical mechanics and quantum mechanics, rapid increment in computer speed and memory, algorithm efficiency and steady improvements in force field development. Modeling and simulation of chemical and biological systems is a truly multidisciplinary challenge. Schlich writes, biologists describe the cellular picture; chemists fill in the atomic and molecular details; physicists extend these views to the electronic level and the underlying forces; mathematicians analyze and formulate appropriate numerical models and algorithms; and computer scientists and engineers provide the crucial implementational support for running large computer programs on high-speed and extended-communication platforms (Schlich, 2010).

The role of computer science becomes constantly even more important as computer simulations become so precise and accessible that they can support and even substitute real experiments. Although nowadays, modeling of large and complex systems as proteins, nucleic acids, and lipids has been made available, computational intensity still remains as a problem (Gubbins, et.al., 2011). Molecular modeling has a wide range of applications also in various disciplines of engineering sciences, such as material science, chemical engineering, biomedical engineering, etc. Knowledge provided by molecular modeling, is essential for understanding the behavior of nanosystems and it forms the route to the nanosciences and nanotechnology. (Mashaghi, et.al., 2013)

Lipids are organic molecules that include fats, waxes, sterols, fat-soluble vitamins, monoglycerides, diglycerides, triglycerides, phospholipids, and others. Phospholipids - the dominant lipids in biomembranes are molecules with hydrophobic tails and hydrophilic head groups. The head groups can be charged (positively or negatively) or neutral. Thanks to hydrophobic nature of their tails, in solution lipids can self-assemble into different nano-structures such as bilayers, liposomes, micelles or reversed micelles. (Alberts et al., 1994) The main lipid biological functions are energy storage, signaling, and acting as "building blocks" of cell membranes.(Fahy et.al.,2009, Subramaniam et.al.,2011) Lipids have many applications in cosmetic and food industries as well as in nanotechnology.(Mashaghi et.al., 2013) Recent researches show that lipids work as nanocarriers in drug delivery using lipid nanotechnology for cancer and tumor treatment. (Selvamuthukumar and Velmurugan, 2012). There are also studies of drug transdermal diffusion by modeling molecule transfer through lipid bilayer - compounds that are soluble in oil, can be delivered by the intercellular lipid layer. (Rim et.al., 2009)

This paper focuses on software aided methodology of molecular dynamics using classical modeling cycle that can be used for investigation of complex lipid system molecular dynamics such as the cationic amphiphilic lipid type compound 1,1'-{[3,5-bis(dodecyloxycarbonyl)-4-phenyl-1,4-dihydropyridin-2,6-diyl]dimethylene} bispyridinium dibromide (1,4-DHP lipid). This work is continuation of our earlier studies, where it was confirmed that 1,4-DHP lipid has the gene transfection activity. (Liepina et al., 2011)

Materials and methods

In this paper we focus on two main research objectives:

- to summarize and classify molecular dynamics software that can be used for 1,4-DHP lipid system molecular dynamics;
- to provide the worked out software aided methodology for molecular dynamics using classical modeling cycle.

As it is not possible to cover the whole range of accessible software for molecular dynamics, authors' selected and listed software choice is subjective and based on scientific analysis and evaluation of a number of information sources and reports connected with molecular modeling and dynamics and also on the authors' reflection and research experience.

Molecular mechanics. Molecular mechanics uses classical mechanics to study small molecules as well as large biological systems or material assemblies with many thousands to millions of atoms. All-atomistic molecular mechanics methods have the following properties: each atom is simulated as a single particle; each particle is assigned a radius, polarizability, and a constant net charge; bonded interactions are treated as "springs" with an equilibrium distances and angles equal to the experimental or calculated values. Main branches of molecular mechanics applications are energy minimization and molecular dynamics.

In molecular dynamics, trajectories of the molecular systems that consist of atoms are generated by integrating Newton's laws of motion. The result is a trajectory that specifies how the positions and velocities of the atom in the system change in time. The trajectory of *i-th* atom is obtained by solving the differential equation embodies in Newton's second law (*F=ma*):

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}, (i = 1 \ldots n) \tag{1}$$

where m_i mass of atom *i*;

x_i – one coordinate of atom *i*;

F_{x_i} – force that acts on the atom *i* in x_i direction.

Equation (1) describes the motion of an atom *i* of mass m_i along one coordinate x_i with force F_{x_i}. The force that acts on the atoms depends on its positions relative to the other atoms. Here the motion is often very difficult, sometimes impossible, to describe analytically, due to the coupled nature of the atoms' motions. The force on each atom will change whenever the atom change its position, or whenever any of the other atoms with which is interacts changes position. These interactions are described by the force field. (Leach, 2001)

Force field. The usage of the term "force field" in molecular modeling differs from the standard usage in physics. Here it is a system of potential energy functions rather than the gradient of potential, as defined in physics. Molecular mechanics ignore the electronic motions that are used in quantum mechanical methods and calculate the energy of system as a function of atoms nuclear positions only. This gives the opportunities to perform the calculations on systems containing significantly larger number of atoms. Molecular mechanics is based upon a rather simple approximation model of the atom interactions within system that includes such processes as bond stretching, opening and closing of angles and rotation about single bonds. (Leach, 2001) A force field is built up from two distinct components to describe the interaction between atoms:

- the set of equations, also called the potential functions, used to generate the potential energies and their derivatives, the forces;
- the parameters that are used in this set of equations.

One functional form for such a force field that can be used to model single molecules or systems of atoms is:

$$V(r_N) = \sum_{\text{bonds}} k_b (l - l_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \\ + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \eta_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\eta_0 r_{ij}} \right\} \tag{2}$$

$V(r_N)$ denotes the potential energy, that is function of the positions r (usually is three dimensional Cartesian space) of N atoms. The first term in equation (2) models interactions between pairs of bonded atoms and is expressed by harmonic potential that gives the increase in energy as the bond length l_i deviates from the equilibrium value l_0. The second term in (2) is a summation over all valence angles that are formed between

three atoms and expressed also in the form of harmonic potential. The third term in (2) is rotational potential that models how the energy changes when torsion angles change. The fourth contribution is the non-bonded term. This is calculated between all pairs of atoms (*i* and *j*) that are in different molecules or are separated in one molecule by at least with three bonds. Non-bonded terms are usually modeled with Lennard-Jones potential for van der Waals interactions and Coulomb potential for electrostatic interactions.

From the potential energy function, mathematical equations, is obtained empirical force field, equations and parameters that relate chemical structure and conformation to energy. All force fields are based on numerous approximations and derived from different types of experimental data. Therefore they are called empirical. There are three main types of force fields: all atom - parameters provided for every atom within the system, united atom – some atoms are excluded, coarse grained - an abstract representation of molecules are created by grouping several atoms into one unit. In the literature (Mackerell, 2004), often is proposed classification as Class I (or classical) Class II and other specific force field models. Some force fields are specifically developed for some biological molecule types and system sizes. AMBER (Case et.al., 2012), OPLS, CHARMM (Vanommeslaeghe et.al.,2010) and GROMOS (Hess et al., 2008) have been developed primarily for molecular dynamics of macromolecules. These are classical, all atom force fields and developers provide information about their force field parameterization strategy. But as they are different, parameters from one force field cannot usually be used in another force field. (Leach, 2011, Ramachandran et.al., 2008, Schlich, 2010, Griebel et.al.,2007)

Molecular dynamics simulation. Due to the complexity of the force field, equation of motion (1) is integrated using finite difference method – integration is broken down into many small stages, each separated in time by a fixed time δt . The wide variety of integration schemes are available - *Verlet algorithm, velocity Verlet method, Beeman's algorithm, Gear predictor – corrector algorithm, etc.* (Leach, 2001)Various factors should be taken account before deciding which method is the most appropriate. It is clear that large computational effort is required for complex system integration, but for best methods high-speed computation is as valued as trajectory precision.

Before running the molecular dynamics simulation, initial configuration of the system should be established. The initial configuration can be obtained from the experimental data or from the theoretical model using energy minimization techniques. It is necessary to assign initial velocities for particles in the system and that can be done by randomly selecting initial velocities from Maxwell-Boltzmann distribution at the temperature of interest. Molecular dynamics is performed in the constant microcanonical ensemble, depending on which state variables are kept fixed – energy E, volume V, temperature T, number of particles N. Two most common alternative ensembles from the traditional constant NVE (number of particles, volume and energy) and constant NVEP (number of particles, volume, energy and pressure) are constant NVT (number of particles, volume and temperature) and constant NPT (number of particles, pressure and temperature) ensembles. (Ramachandran et.al., 2008)

Results and discussion

Software aided methodology of molecular dynamics. The fact that molecular modeling includes a whole set of theoretical and computational methods that are used to investigate and simulate behavior of molecular systems, should be taken into account when talking about proper software tools. There are available many self-sufficient software tools that can be used by molecular modelers from beginners to advanced scientists. Some of them include many methods covering a wide range, while others are concentrating on a very specific range or even a single method. By authors' vision, there is no strict classification for molecular modeling software provided in the literature. Of course all software highly depends on the developers. Some of them have made more successful implementations of methods than others however some of them historically or by other assumptions are more popular among users. Authors propose to systematize molecular modeling software by some qualitative features*:*

- *Functionality* – what methods that are implemented in software, e.g., molecular model building, visualization, energy minimization, molecular dynamics, stochastic molecular dynamics, protein folding, protein structure prediction, etc.
- *Supported biomolecules* – what biological systems are supported in the software, e.g., nucleotides, proteins, lipids or saccharides.
- *Type of graphical interface* – e.g., graphical user interface, command line, batch interface.
- *Type of license* – e.g., commercial software, GNU general public licence, open source, etc.
- *Supported operating system (OS)* – various Windows OS, Macintosh OS, Linux OS, Unix based.

In further analysis only software suitable for 1,4-DHP lipid molecular dynamics process will be discussed. Molecular dynamics simulation as a computer modeling method complies with the assumptions of the mathematical modeling cycle. Classical mathematical modeling competences and cycle more detailed is described by Duka (Duka, 2012). Molecular dynamics modeling cycle consists of five steps: real world problem,

mechanical molecular model, computer model, molecular dynamics simulation, conclusions. These five steps can be repeated cyclically until best model for real world problem representation is chosen. (Fig. 1)

The beginning of molecular dynamics simulation cycle is the *model building part*. From the real world problem the mechanical molecular model and after then also computer model is formed. Mechanical molecular model follows from the theoretical background of molecular dynamics method and molecular modeling theory. As a molecular modeling is an interdisciplinary field, model in molecular dynamics encompass not only mathematical but also physical, chemical, biological theories. Then mechanical molecular model is transferred in computer model development. Simple computer model, that is suitable for molecular dynamics simulation, is data file with descriptions of atom types, atomic coordinates in three dimensional Cartesian coordinate space, atomic connectivity, etc. These computer models are often prepared in the Protein Data Bank (pdb) file format or XYZ file format. There are many ways how to create initial computer models of molecular system. Model building can be carried out manually, by hand, but software assistance as, for example, 3D graphical molecular builders and editors can be used. Most of the model building software that handle large molecular system building are commercial and supported also with graphical user interface, e.g., MOE (Molecular Operating Environment, 2012), MacroModel (MacroModel, 2012). Detailed description of software list in (Table 1)

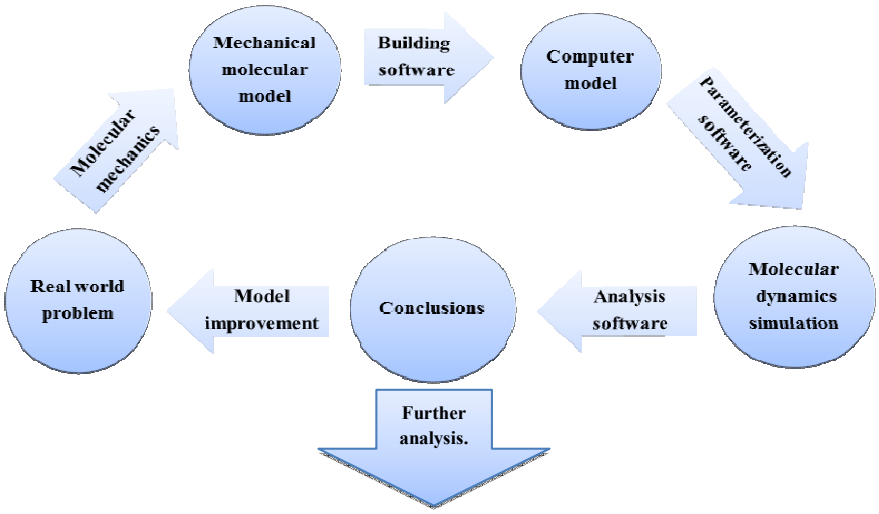


Fig. 1. **Modeling cycle of molecular dynamics simulation.**

After creation of molecular system computer model, it should be prepared for molecular dynamics simulation. In preparation process initial structure is solvated in periodic box, octahedron of, for example, water and then constructed force field files for this system. After such preparation, from initial system (usually one file) several data files are created where the information about topology, trajectory and simulation parameters is going to be kept. Different software tools are used for preparing input files for the simulation programs. For standard molecular dynamics simulation, common known molecular dynamics software developers have grown together with their provided force fields and parameters. The transfer between softwares and force fields is not recommended during single simulation process, therefor preparation, parameterization and simulation of molecular system is usually done using tools from the same developer. Must be noted that for standard molecular structures force field parameters are assigned from known databases, but for non-standard systems even for common used GROMOS, AMBER, and CHARMM force fields, this derivation often takes the form of various quantum mechanical calculations. Also for this reason, automated tools are greatly preferred. For each force field, there are methodologies or software programs for assigning parameters for molecular structures, compatible with various force fields. Using AMBER software for molecular dynamics, Antechamber and Leap packages from Amber Tools software (Case et.al., 2012) prepares the molecular systems and applies all atom AMBER force field to the molecule. CGenFF can be used for generalized force field assignment for CHARMM (Brooks, B. R., et.al., 2009). For molecular dynamics simulation in GROMOS87/GROMOS96 force fields with GROMACS molecular dynamics software, also PRODRG 2.5 (Schüttelkopf and Aalten, 2004) ATB (Automated Topology Builder) (Malde et.al., 2011) web server online services can be used as an automated servers for topology generation. Molecular dynamics simulations in these force fields can be calculated also using other softwares, such as Abalone, NAMD (Phillips et.al., 2005), Ascalph, Maestro (Maestro, 2012), MOE, Desmond (Bowers, 2006). Some of them are more advanced than another and direct graphical dynamics can be very illustrative for demonstrations. But when the research work comes to large biological systems, then the fact that molecular dynamics calculations are time and resource demanding must be noted, and then molecular dynamics

simulations are carried out on a remote Unix based servers in computer centres or laboratories using command line and batch interface. Detailed description of software list in (Table 1)

Following the modelling cycle principles (Duka, 2012), after a molecular dynamics simulation, result analysis should be carried out for making decent conclusions about the behaviour and structure of investigated molecular system. Analysis, like mean energy, density of the system, RMS difference between two structures etc., can be performed manually using different scripts, that reads and represents the molecular dynamics information from large number of data files or specific analysis tools for different purposes can be used. For graphical representations of molecular system and obtained dynamics – systems trajectory, a list of visualization softwares can be found, Vega ZZ (Pedretti, 2004),VMD (Humphrey et.al., 1996), RASMOL (Sayle and Milner-White, 1995, Bernstein, 2000), MOIL (West et.al., 2007), MOE, Maestro, MacroModel, Abalone, Ascalph. Most of the tools for molecular dynamics come together with functions for static or dynamical graphical representations and analysis possibilities, such as MOE, Maestro, MacroModel, some of these tools are only for representation, like Rasmol. Detailed description of software list in (Table 1)

Very important part of molecular dynamics simulation is verification process. Thanks to the many research groups that have been working and developing molecular dynamics, this method has become as independent research method with high precision. However, for every computer simulation result testing should be done either by literature review or comparison with experimental data.

Advantages of modeling is in its possibilities - going beyond visible, seeing further and discovering more using simplified model and computer technologies. The role of molecular dynamics is highly valued especially for the possibility to notice and prove regularities that could not have been seen with any other microscopic methods. Analysis, hypothesis testing and prediction come as another important part of molecular dynamics cycle.

Molecular modeling of 1,4-DHP lipid. This study of molecular dynamics simulation is offered as an example of successful computer experiment. Molecular model of one 1,4-DHP lipid was built using MOE software and lipid bilayer consisting of 72 molecules of 1,4-DHP-lipid was created manually using self-written coordinate transformation script written in programming language C++. With software package Leap from Amber Tools 8.0 1,4-DHP lipid system initially transformed into a periodic lipid bilayer-water box, with 10 Å water over the solute and with a small amount of excessive water on the lipid edges to ensure the mobility of lipid molecules. There were 72 DHP-lipid molecules, 144 counterions of chlorine ions and 4401 water molecules in the system DHP-lipid-water box. The total number of atoms in the system is 22491.

After the 1,4-DHP-lipid-water box were subjected to molecular dynamics, (AMBER 8.0 (f99) force field, version 8.0 [14-16], NTP protocol (constant number of particles, constant temperature, constant pressure). The temperature was increased gradually from T = 10 K by step of 10 degrees till 300 K. Calculations were performed for 326212 ps. Molecular dynamics simulations were started from the temperature T = 10 K and risen gradually till 300 K. Analysis of molecular dynamics results were proceeded with VMD, visualization were made with RASMOL.

The1,4-DHP-lipid-water box system kept the initial bilayer organization at the beginning of the MD simulation, but during MD run lipid molecules squeezed from one bilayer into another, finally forming worm-like micellae. Results of 1,4-DHP lipid MD simulation show that 1,4-DHP lipid in natural conditions does not form a lipid bilayer, but one of its structures is a tubular worm-like micellaes. We could expect that such the micellaes are capable to form a functional structure for the DNA transfection. Result was confirmed with the studies electron microscopy showing extended, worm-like structures as one of the possible 1,4-DHP lipid structures.

Table 1

List of software for performing complete modeling cycle of lipid molecular dynamics simulation

Name	MB(1)	MD(2)	GR(3)	L(4)	Developer	Supported OS	Interface	License
Abalone	+	+	+		Agile Molecule	Windows XP	Graphical user interface	Commercial
AMBER, AMBER Tools	+	+		+	AMBER developer project	Linux, Various Unix workstations	Command line, Batch interface	Commercial
Ascalaph	+	+	+		Agile Molecule	Windows 2K/XP	Graphical user interface	GNU General Public License

Atomistix ToolKit	+	+	+	+	Quantum Wise	Windows XP/Vista/7, Linux	Graphical user interface, Command line	Commercial
CGenFF	+				Harvard University	Linux and Various UNIX workstations	Command line. Batch interface	Commercial, Open source
CHARMM		+		+	Harvard University	Linux and Various UNIX workstations	Command line. Batch interface	Commercial, Open source
Desmond		+			D.E.Shaw Research	Linux and Various UNIX workstations	Command line Batch interface	GNU General Public License, Open Source
GROMACS		+			GROMACS project group	Solaris, Linux, OS-X, Windows Various UNIX workstations	Command line. Batch interface	GNU General Public License Open Source.
MAESTRO MacroModel	+	+	+	+	Schrödinger	Windows XP/Vista/7, Linux, Mac OS X	Graphical user interface, Command line	Commercial
MOE	+	+	+	+	Chemical Computing Group	Windows XP/Vista/7, Linux, Mac OS X	Graphical user interface, Command line	Commercial
MOIL			+		MOIL team	Windows, MacOsX, Linux (Fedora)	Graphical user interface, Command line	GNU General Public License, Open Source
Rasmol			+		Roger A. Sayle, Herbert J.Bernstein	Windows, MacOS, UNIX, VMS systems, etc	Graphical user interface, command line.	GNU General Public License, RASLIC license
VMD NAMD		+	+	+	University of Illinois at Urbana-Champaign	MacOS X, Unix, or Windows	Graphical user interface, Command line, Batch interface	GNU General Public License, Open Source
Vega ZZ		+	+	+	Drug Design Laboratory	Windows Linux and Various UNIX workstations	Graphical user interface, Command line	GNU General Public License, Open Source

(1) MB – Software for molecular model building
(2) MD – Software for molecular dynamics simulation
(3) GR – Software for graphical representations of molecular systems
(4) L – Applicable for lipid analysis

Conclusion

We summarized and systematized the molecular dynamics simulation process, and provided a list of software tools that can be successfully used for different purposes in different steps of molecular dynamics simulation. We showed that molecular dynamics simulation as a computer modeling method complies with the assumptions of the mathematical modeling cycle. That was proved with the successful molecular dynamics studies of 1,4-DHP lipid system. The cycle of 1,4-DHP lipid system modeling was accomplished with the verification of results, that is excellent result although it opens a perspectives for further analysis of this system, while the last step of modeling cycle - analysis and prediction, is still neglected. Further studies promise challenges in the field of availability of the molecular modeling software, while 1,4-DHP lipid system deviated from the standard lipid bilayer structure and formed tubular, worm-like structure. Tubular structure lipid systems are investigated less with molecular modeling methods and software tools.

Acknowledgements

Calculations were performed on computers of the Gdansk Academic Computer Center TASK.

References

Alberts, B, Bray, D, Lewis, J, et al., 1994. Molecular Biology of the Cell. The Lipid Bilayer. *3rd edition. New York: Garland Science*.

Bernstein, H.J., 2000. Recent changes to RasMol, recombining the variants. Trends in Biochemical Sciences (TIBS) , 25(9), pp. 453-455

Bowers,K.J., 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters, *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, November 11–17.

Brooks, B.R., et.al., 2009. CHARMM: The Biomolecular simulation Program. J. Comp. Chem. 30, pp.1545-1615

Case, D.A., Darden, T.A., et.al., 2012, AMBER 12, University of California, San Francisco.

Hess, B., et al., 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. Journal of Chemical Theory and Computation, 4(3), pp.435-447.

Humphrey, W., et.al., 1996. VMD- Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14, pp.33-38. DOI:10.1016/0263-7855(96)00018-5

Duka, V. and Zeidmane, A., 2012. Importance of mathematical modelling skills in engineering education for master and doctoral students of Latvia University of Agriculture. 2012 15th International Conference on Interactive Collaborative Learning, ICL 2012, Conference Paper, ISBN: 978-146732427-4, DOI: 10.1109/ICL.2012.6402192

Fahy, E, Subramaniam, S, et.al., 2009. Update of the LIPID MAPS comprehensive classification system for lipids. Journal of Lipid Research 50 (Supplement): S9–S14. doi:10.1194/jlr.R800095-JLR200. PMC 2674711. PMID 19098281.

Leach, A.R., 2001. Molecular modeling principles and applications. *Pearson Educated Limited, 2nd edition.* ISBN:987-0-582-38210-7

Liepina, I. et al., 2011. 1,4-DHP-lipid parameters and rod like micellae. Journal of Biophysical Chemistry, 2(4), pp.386–394.

Griebel, M., Knapek, S., Zumbusch, G.W., 2007. Numerical simulation in molecular dynamics, Springer Berlin.

Gubbins, K.E., Liu, Y.C., Moore, J.D., and Palmer, J.C. 2011. The role of molecular modeling in confined systems: impact and prospects. *Physical chemistry chemical physics : PCCP*, *13*(1), pp.58–85. doi:10.1039/c0cp01475c

Mackerell, A.D., 2004. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13), pp.1584–604.

MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2012.

Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012.

Malde, AK, et.al., 2011. An Automated force field Topology Builder (ATB) and repository: version 1.0. Journal of Chemical Theory and Computation, 7(12), pp.4026-4037. DOI: 10.1021/ct200196m

Mashaghi, S., Jadidi, T., Koenderink, G., and Mashaghi, A., 2013. Lipid Nanotechnology. Int. J. Mol. Sci. 14, 2: pp.4242-4282.

Molecular Operating Environment (MOE), 2012.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.

Pedretti, A., et.al., 2004, Vega - An Open Platform To Develop Chemo-Bio-Informatics Applications, Using Plug-In Architecture And Script Programming, *J.C.A.M.D.*,18, 167-173.

Phillips, J.C., et.al., 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:pp.1781-1802.

Ramachandran, K.I., Deepa, G. and Krishnan Namboori, P.K., 2008. Computational Chemistry and Molecular Modeling Principles and applications. *Springer-Verlag GmbH*.

Rim, E. , Pinsky, P.M. , and Osdol, W.W. , 2009. Multiscale modeling framework of transdermal drug delivery. *Annals of biomedical engineering*, 37(6), pp. 1217–1229.

Sayle, R. and Milner-White, E.J., 1995. RasMol: Bio- molecular graphics for all. Trends in Biochemical Sciences (TIBS), 20,pp. 374. doi:10.1016/S0968-0004(00)89080-5Selvamuthukumar, S., and Velmurugan, R., 2012. Nanostructured Lipid Carriers: A potential drug carrier for cancer chemotherapy. *Lipids in health and disease*, 11, p. 159.

Schlich, T., 2010. Molecular Modeling and Simulation: An Interdisciplinary Guide. *SpringerLink, An Interdisciplinary Guide, Interdisciplinary Applied Mathematics*, Vol. 21,2nd Edition., XLV, 723 p., ISBN 978-1-4419-6351-2

Schüttelkopf, A.W, and Van Aalten, D.M.F., 2004. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D60, pp.1355–1363.

Subramaniam, S, Fahy, E, et.al., 2011. Bioinformatics and Systems Biology of the Lipidome. Chemical Reviews 111 (10): pp.6452–6490. doi:10.1021/cr200295k. PMID 21939287.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. and Mackerell, A.D., 2010. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem., 31: pp.671–690. DOI: 10.1002/jcc.21367West, A. M. A., et.al., 2007, MOIL, Journal of Chemical Physics 126, 14.

The role of ICT in the supply chain resilience

Peter Mensah, Yuri Merkuryev

Department of Modeling and Simulation, Riga Technical University, 1 Kalku Street, LV-1658 Riga, Latvia
E-mail: p_mensah@hotmail.com, merkur@itl.rtu.lv

Abstract: *The supply chain, with so many definitions consisting of overlapping terminology and meanings, has evolved greatly over the past 50 years from the traditional form where big and powerful companies used to be wholly and solely responsible for supplies, manufacturing and distribution to the modern innovative companies that are actually outsourcing almost all the processes in the supply chain. As many research programs have also shown that modern supply chains are at greater risks than their supply chain managers recognise, supply chain vulnerability has become a very significant issue for many companies. These risks including natural disasters, terrorism, cyber attacks, credit crunch and many more could yield to a drastic loss in productivity, revenue, competitive advantage, profitability etc, if not managed appropriately. Should in case one of the risks occurs and therefore deforming the supply chain, the possibility of the supply chain returning back to its original state is a concern and this is where developing a resilient supply chain is of great importance. Hence, resilience is now vital in the supply chain management since the numbers of threats that can undermine a supply chain are now greater. With cutting-edge technology on the increase, ICT would play a vital role in reducing risk, increasing management efficiency resulting in a more resilient supply chain. The* ***aim of this paper*** *is therefore to analyze the definition of the supply chain, investigate the sources of the supply chain risk and elaborate on the role of ICT in developing a more resilient supply chain.*

Keywords: supply chain management, supply chain risks, resilient supply chain, Information and Communication Technology.

Introduction

This paper discusses a theoretical approach in developing a supply chain resilient strategy with Information and Communication Technologies (ICT) playing an important role that would help organizations to bounce back after deformation in any area along the supply chain.

According to the research findings of the 3rd Annual Survey of the Business Continuity Institute in Supply Chain Resilience in 2011, where more than 550 organizations from over 60 countries were surveyed, 'Supply chain incidents led to a loss of productivity for almost half of businesses along with increased cost of working (38%) and loss of revenue (32%)' (The Business Continuity Institute, 2011).

These incidents were due to natural disasters, terrorism, cyber-attacks, credit crunch and many more, and as they are unpredictable it is now important for organizations to develop a more resilient supply chain.

The above has motivated the authors to conduct a research and recommend certain strategies in the area of the supply chain resilience.

The **research methods** are mainly qualitative where various relevant literatures, journals and other scientific publications as well as the authors' professional lecturing experiences in the field of management and modelling and simulation were exploited.

On the other hand, quantitative secondary resources were also used to portray the degree of the loss of productivity due to supply chain incidents.

The **objectives of the research** is to develop a supply chain resilient strategy with ICT playing an important role along the upper stream and lower stream that will help organizations to bounce back after deformation on any area along the supply chain.

The paper is divided into four chapters namely, the supply chain, the supply chain management, the supply chain resilience and the role of ICT in the supply chain resilience.

Various definitions of the Supply Chain and the Supply Chain Management are analysed. The authors introduce a new and more realistic definition for the supply chain as the supply chain management consists of many definitions with overlapping terminology and meanings (Croom et al., 2000). In addition, various strategies and tools are also recommended in order to implement an effective supply chain management.

The next chapter which is the Supply Chain Resilience investigates the issues facing today's organizations and how they affect the supply chain, and in case of any deformation along the upper stream and or lower stream, it portrays the possibilities of organizations bouncing back to normal through certain recommended strategies given by the authors.

The last chapter deals with the role of ICT in the Supply Chain Resilience as it speeds up information and financial flows between the upstream and downstream of the supply chain resulting in improved services, reduced logistics costs and faster communication between customers and their suppliers.

Defining the Supply Chain

It is evident that over the past 50 years the Supply Chain has been defined differently where they consist of overlapping terminologies and meanings by different authors. Although these definitions general cover most of the sectors of the supply chain, a more appropriate definition is yet to be determined. In order to come up with a new definition, the authors of this paper have randomly selected and analyzed ten out of many definitions between 1999 and 2011 given in Table 1 below.

Table 1

Various Definitions of the Supply Chain

Author(s)	Year	Definition of Supply Chain
Lu, D.	2011	...a group of inter-connected participating companies that add value to a stream of transformed inputs from their source of origin to the end products or services that are demanded by the designated end-customers...
Pienaar, W.	2009	...a general description of the process integration involving organizations to transform raw materials into finished goods and to transport them to the end-user...
Bridgefield Group	2006	...a connected set of resources and processes that starts with the raw materials sourcing and expands through the delivery of finished goods to the end consumer...
S, Cholette	2011	...a sequenced network of facilities and activities that support the production and delivery of a good or service...
Sunil, C., Meindl, P.	2004	...consists of all parties involved, directly or indirectly, in fulfilling a customer request. The supply chain not only includes the manufacturer and suppliers, but also transporters, warehouses, retailers, and customers themselves. Within each organization, such as manufacturer, the supply chain includes all functions involved in receiving and filling a customer request...
Croker, J.	2003	...a total flow of materials, information and cash through a business network, all the way from the suppliers' suppliers to the customers' customers...
Tecc.com.au	2002	...a chain starting with raw materials and finishing with the sale of the finished good...
Ayers, J. B.	2001	...life cycle processes involving physical goods, information, and financial flows whose objective is to satisfy end consumer requisites with goods and services from diverse, connected suppliers...
Little, A.	1999	...the combined and coordinated flows of goods from origin to final destination, also the information flows that are linked with it...
Beamon B.	1998	...a structured manufacturing process wherein raw materials are transformed into finished goods, then delivered to end customers...

From Table 1 above, most of the definitions are similar and too general and basically deals with processes that start with raw materials and transformed into finished goods that will eventually reach the customers. However, three of the definitions (Sunil, 2004; Croker, 2003; Ayers, 2001) included materials, information and financial flows in their definition that plays a crucial role in the supply chain.

After careful analysis, the authors of this paper have developed a new supply chain definition which is given below:

'*Supply chain is a sequenced network of business partners involved in production processes that convert raw materials into finished goods or services in order to satisfy the consumers' demand.'*

The Supply Chain Management

Just like the Supply Chain, the Supply Chain Management (SCM) has numerous definitions.

One of the most recent definitions of the SCM according to CSCMP (Council of Supply Chain Management Professionals) states that 'Supply Chain Management encompasses the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third party service providers, and customers (CSCMP, 2013). In essence, supply chain management integrates supply and demand management within and across companies'. Furthermore, Supply Chain Management is aimed at examining and managing Supply Chain networks (Janvier, 2012).

To simplify the above, the authors would describe the SCM as the Business Management of the activities of the Supply Chain aimed at satisfying the demand of the end consumer. In order for an effective SCM to be implemented, three levels of decision making namely planning, strategy and operations have to be considered.

Planning:

What is to be done and how it should be done are questions if answered to correctly; then the organization is on the right path. As a matter of fact, planning consists of defining the organization's goals, establishing a strategy to achieve the necessary goals, and on the other hand integrating and coordinating activities through proper arrangements. All of these lead to the development of an appropriate strategy for the supply chain.

The strategic decision making:

After planning, the supply chain has to be designed, crafted and developed for the production of particular goods or services. As top managers are responsible for decision making about the mission and direction of the organization, and establishing policies that affect all organizational members, they will have to take the responsibilities of crafting the appropriate cost effective supply chain that is responsive to customers' demand, whilst facilitating product development, manufacturing and logistics.

Operations:

The application of the transformation process where organizations add value by converting inputs into outputs along the supply chain is known as operations. From a general perspective point of view, it is organized by determining what tasks to be done and by whom; how the tasks are to be grouped; who reports to whom and where decisions are to be made.

Supply Chain Resilience

Many research programs have also shown that modern supply chains are at greater risks than their supply chain managers recognise. 'In today's uncertain and turbulent markets, supply chain vulnerability has become an issue of significance for many companies' and appropriate research on resilient supply chain are yet to be conducted (Christopher, 2004; Peck, 2004).

The risks including natural disasters, terrorism, cyber attacks, credit crunch shrinking product lifecycles, volatile and unpredictable markets and many more, could yield to a drastic loss in productivity, revenue, competitive advantage, profitability etc, if not managed appropriately.

Should in case one of the risks occurs and therefore deforming the supply chain, the possibility of the supply chain returning back to its original state is a concern and this is where developing a resilient supply chain is of great importance.

As the numbers of threats that can undermine a supply chain are now greater, organizations are facing greater challenges in managing risks (Sheffi, 2005). Hence, resilience is now vital in the supply chain management.

According to the research findings of the Business Continuity Institute in Supply Chain Resilience in 2011, where more than 550 organizations from over 60 countries were surveyed, 'Supply chain incidents led to a loss of productivity for almost half of businesses along with increased cost of working (38%) and loss of revenue (32%)'.

Although 'resilience' could be defined as the ability of a substance to go back to its original state or form after deformation, the Supply Chain Resilience is still a new area of management to be explored as the possibility of it returning back to its original form after deformation is still 'theoretical', and many organisations still lack the awareness that it is necessary to take into consideration a resilient supply chain as part of their strategy when developing their risk and business continuity management.

Developing a Resilient Supply Chain:

Historically, immediate solutions to disrupted supply chains like the outbreak of the SARS or the Hurricane Sandy that affected 24 states in the USA in October last year that brought many businesses and production into a halt due to flooding streets, subway lines and tunnels, and power failure in some cities especially New York, was not possible and might not be possible in the future if organizations lack proper planning.

However, organizations with better planning, strategy and operations tend to have a way out and competitive advantages over their rivals after disruptions in the short run.

According to the authors of this article, organizations would be able to develop a more resilient supply chain by planning and implementing six sigma practices, lean production, flexibility and a strong corporate culture. Consequently, these organizations would have the capabilities to speed up the process of bouncing back after deformation on any part along the supply chain. The authors have tabularised these strategies in table 2 below in order to portray their effectiveness and shortcomings.

All of the strategies in Table 2 are significant and could be adopted and practiced in organizations depending on their capabilities.

Due to the limitation of this paper, the authors will be focusing on the theoretical analysis of the six sigma strategy only.

Table 2

Resilient Strategies (The authors)

Strategy	Advantages	Disadvantages	Examples
Six sigma supply chain	3.4 defects per million activities or opportunities. Stops and prevents problems from happening. Management will be able to solve problems effectively as they have a solid grasp on the problems of their organization. Pay off in the long run.	High cost in investing both time and money into training employees in order to utilize the sigma tools effectively. A long-term method.	Ford General Electric Motorola Allied Signal
Lean production with JIT delivery and low inventory	Minimise waste and inefficiency. Continuous improvement in quality, productivity and responsiveness. Tight control over production process. Shortening product development cycles.	Necessitates rapid and frequent flow of goods and information Entails close relationship with suppliers	Kellogg's
Increasing SC flexibility	Better respond to a change in demand Capabilities in reallocating resources when needed Developing good relationship with suppliers	Basically no disadvantages except for the fact that employees have to be trained which is an increase in costs	Intel
Developing a strong corporate culture	Employees well informed about the organization activities through continuous communication. Empowering employees to make quick decisions. Quick recovery after disruptions.	Might create dysfunctional conflict among employees if they are not well informed.	Toyota Apple

Six Sigma:

The Six Sigma, which is currently used by a number of successful companies, was originally started by the Motorola Corporation during the 1980s. It was designed to solve business problems like declining sales and poor performance among employees.

The Six Sigma is also a highly potent statistics-based methodology that helps to eliminate defects, optimize processes, and produce significant financial results across the organization. This is evident in Ford as it has gained tremendous success by applying Six Sigma to its supply chain processes (Moore, 2002).

Sigma tools like the Fishbone Diagram and the FMEA (Failure Modes and Effects Analysis) are the most common ones used today as companies will be able to make amendments in their production processes (Aveta Business Institute, 2013). Furthermore, according to Graeme Knowles, all projects should be conducted through the DMAIC (Define-Measure-Analyse-Improve -Control) methodology as shown in Fig. 1 below.

With reference to Fig. 1, in the Define process step, the organization needs to know what it is seeking to improve and should therefore define its opportunity correctly as failing to do so will lead to a catastrophic output. After defining the process, the next step is 'Measure' where current performances should be analysed and any variability and its sources identified before moving onto the 'Analyse' step where the relationship among the key variables are analysed through cause and effect analysis, called Ishikawa, and prioritise for actions (Christopher, 2004; Rutherford, 2004).

The next process step 'Improve', involves re-engineering the process and implementing its solution, where predicted results are also tested.

'Control' comes next with its key variables being controlled and the performance monitored. Finally, the benefits achieved should be measured and if the benefits coincide with the advantages of the six sigma strategy in Table 2, then the organization is on the brink of achieving a supply chain resilient.

Implementing the six sigma also needs a good leadership which is provided by a team of champions namely, senior champion at the corporate level; deployment champion at the unit level and project champion at the department level all maintained by a team of experts.

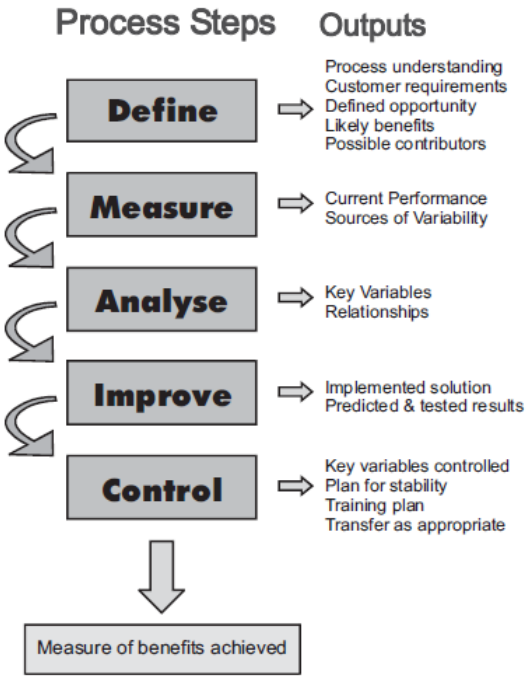


Fig. 1. **The Six Sigma DMAIC Process and Key Outputs** (Canales et al., 2005).

The experts are referred to as Black Belts and Master Black Belts. The Master Black Belts provide mentoring, training and expert support to the Black Belts whilst the Black Belts usually work full time on projects at process level to solve critical problems and achieve bottom-line results (Canales et al., 2005).

ICT and the Supply Chain Resilience

The role of ICT in the supply chain is not only significant because it speeds up information and financial flow between the upstream and downstream of the supply chain resulting in improved services, reduced logistics costs and faster communication between customers and their suppliers, but it also helps the organization to make accurate and quick decisions after deformation of its supply chain that will eventually bounce it back to normal activities.

How does the ICT help to make the activities of an organization bounce back after a supply chain disruption?

The recommended strategies and tools in the chapter of the supply chain resilience namely, six sigma practices, lean production, flexibility and a strong corporate culture could be linked to a shared ICT infrastructure consisting of six sigma software, Enterprise Resource Planning (ERP) in order to facilitate lean production, and Social Intranet Software to develop a strong corporate culture that would engage management and employees on the activities of the organization through online communication and collaboration.

According to the Times of India's article, how tech is powering Spencer's supply chain, the company Spencer Retail Ltd, has been able to manage the dynamic markets of the retail industry by investing in a robust ERP solution from SAP in 2008 (The Times of India, 2013). Transactions are managed in the company's warehouse through SAP's Warehouse Management System (WMS) whereas the ERP systems have been linked with the suppliers and visibility to all key metrics are provided. As a result, it has helped the company to manage its perishable supply chain especially from the upstream where they have to deal with local farmers.

Another example is that Intel has identical layouts for machinery and production process in its semiconductor fabrication factories that enables it to switch production among facilities in case of supply chain disruptions and when necessary (Sheffi, 2005).

From another perspective, Modeling and Simulation could be used as an application technique to support supply chain design, management and optimization (Longo, 2012) as any network of supply chain (Kilmov et al., 2010) can be easily represented by a stimulating process.

Furthermore, stimulation can be used as decision support tools in order to improve the supply chain management, reduce risks and vulnerability (Longo, 2012).

Hence, the role of ICT in the supply chain is now vital in structuring and implementing the supply chain, speeding up information flow resulting in better communication, and giving the organization the capabilities and competencies to bounce back through accurate and quick decisions after deformation of its supply chain.

Therefore, the entire above if put into practice would yield to a more resilient supply chain in an organization.

Conclusions

In order to fulfil this task, various resources including journals and other scientific publications were discussed in the paper, and by analyzing the Supply Chain and the Supply Chain Management concepts, it was discovered that there are too many definitions overlapping with authors having various definitions from different perspective. In this case, after thorough analysis, a more appropriate definition of the supply chain was introduced which states that '*Supply chain is a sequenced network of business partners involved in production processes that convert raw materials into finished goods or services in order to satisfy the consumers' demand'*. Three levels of decision making namely planning, strategy and operations were also recommended in order to implement an effective SCM. Deformation in any part of the supply chain due to the risks facing the modern supply chains today than the supply chain managers would actually recognise was discussed in the chapter of the supply chain resilience, and appropriate strategies and tools namely, six sigma practices, lean production, flexibility and a strong corporate culture were recommended. Only the Six sigma strategy was discussed in more detail due to the limitation of this paper. However, it is necessary to conduct a research in the other strategies and tools as they are of equal importance.

The role of ICT in the supply chain resilience was discovered to be very important as it speeds up information and financial flow between the upstream and downstream of the supply chain and helps organizations to make accurate and quick decisions after deformation of its supply chain that will eventually bounce it back to normal activities. In addition, Modelling and Simulation was also recommended as it could be used as an application technique to support supply chain design, management and optimization (Longo, 2012) as any network of supply chain (Kilmov et al., 2010) can be easily represented by a stimulating process.

Finally, although the authors have investigated the topic and given recommendations using a theoretical approach, a practical approach of the role of ICT in the supply chain resilience is yet to be investigated and applied in an industry, most probably the auto industry, where simulation based analysis would be developed portraying deformation and its impact on any part of the supply chain as this would make supply chain managers to be well prepared to deal with any occurrence of hazardous phenomena and its impact along the supply chain.

References

Aveta Business Institute, 2013. Six Sigma. Available at: http://www.sixsigmaonline.org/six-sigma-training-certification-information/articles/six-sigma-tools-advantages-and-disadvantages.html, 10.05.2013.

Ayers, J. B., 2001. APICS Series on Resource Management. *Handbook of Supply Chain Management,* The Saint Lucie Press, Boca Raton, Florida.

Beamon, B., 1998. Supply Chain design and analysis: Models and methods. Available at: http://www.damas.ift.ulaval.ca/~moyaux/coupfouet/beamon98.pdf, 05.01.2013.

Bridgefield Group Erp Supply Chain (SC) glossary, 2006. Bridgefield Group. Available at: http://bridgefieldgroup.com/bridgefieldgroup/glos7.htm#P, 03.01.2013.

Canales, F.D.C., Femat, J.H., Knowles, G., Whicker, L., 2005. A conceptual model for the application of Six Sigma methodologies to supply chain improvement. *International Journal of Logistics Research and Applications*, 8, pp. 51–65.

Cholette, S,. 2011. *Sustainable Supply Chain*, Basics Decision Sciences, San Francisco State University USA.

Christopher, M., Peck, H., 2004. Building the Resilient Supply Chain. *International Journal of Logistics Management*, 15, pp.1-13.

Christopher, M., Rutherford, C., 2004. Creating Supply Chain Resilience through Agile Six Sigma. *Bringing the Process of Strategy to Life,* Critical EYE, pp 24-27.

Croker, J., 2003. The supply chain. Increasing value through process involvement. Available at: http://www.clermiston.com.au/Documents/Supply%20Chain%20Increasing%20Value%20Process%20Imp%20JRC%20Handout%202x.pdf, 05.01.2013.

Croom, S., Giannakis, M., Romano, P., 2000. Supply chain management: an analytical framework for critical literature review. *European Journal of Purchasing and Supply Management,* Warwick Business School, University of Warwick, Coventry, UK, pp. 67-83.

CSCMP (Council of Supply Chain Management Professionals), 2013. Available at: http://www.cscmp.se/CLMSite/CLM_logistik.htm, 10.01. 2013

Janvier-James, A. M., 2012. A New Introduction to Supply Chains and Supply Chain Management. Definitions and Theories Perspective. *International Business Research*, 5, pp. 194-196.

Klimov, R., Merkuryev, Y., Tolujev, J., 2010. A Theoretical Framework for a Simulation-Based Analysis of Supply Chain Risk Management. *Managing Risk in Virtual Enterprise Networks: Implementing Supply Chain Principles,* IGI Global, pp. 162-182.

Knowles, G., 2011. *Six Sigma*, Grame Knowles and Ventus Publishing Aps.

Levy, D. L., 1997. Sloan Management Review Winter. Available at: http://www.personal.psu.edu/users/g/i/gis1/LeanIntlSupplyChain.pdf, 10.01.2013.

Little, A., 1999. A European Supply Chain Survey. Available at: http://www.adlittle.be/insights/studies/pdf/european_supply_chain_survey.pdf, 05.01.2013.

Longo, F., 2012. Sustainable supply chain design: an application example in local business retail. *Transactions of the Society for Modelling and Simulation International.* Available at: http://intl-sim.sagepub.com/content/88/12/1484.abstract 10.01.2013.

Lu, D., 2011. Fundamentals of supply chain. Dawel Lu and Ventus Publishing Aps. Available at: http://www.tutorsindia.com/document/management-organisation/fundamentals-of-supply-chain-management.pdf, 05.01.2013.

Moore, K. G., 2002. Six Sigma: Driving Supply Chain Success at Ford. *Supply Chain Management Review,* 6, ILL, pp. 38-43.

Pienaar, W,.2009. *Introduction to Business Logistics,* Southern Africa, Oxford University.

Sheffi, Y., 2005. Building a resilient SC. *Supply chain strategy*, Harvard Business Review, October, 1.

Sunil, C., Meindl, P., 2004. *Supply chain management,* Upper Saddle River, Pearson Prentice Hall.

Tecc.com.au., 2002. Available at: http://www.tecc.com.au/tecc/guide/glossary.asp?letter=S, 05.01.2013.

The Business Continuity Institute, 2011. *Supply Chain Resilience 3rd Annual Survey,* The Business Continuity Institute, Caversham, UK.

The Times of India, 2013. How tech is powering Spencer's supply chain TNN. Available at: http://timesofindia.indiatimes.com/tech/enterprise-it/strategy/How-tech-is-powering-Spencers-supply-chain/articleshow/18063191.cms, 17.01. 2013.

ICT FOR RURAL DEVELOPMENT AND AGRICULTURE

Development of robot manipulator and motion control using inverse kinematics for robotized vegetable harvesting

Armands Kviesis, Vitalijs Osadcuks

Latvia University of Agriculture, Liela Str.2, Jelgava, LV-3001, Latvia
armands.kviesis@inbox.lv, vitalijs.osadcuks@llu.lv

Abstract: *Robot manipulators can be mentioned as one of technological developments in agricultural sector, which can be used in fruit, vegetable harvesting. It is important to do the harvesting very precisely. Precision does not start at the point where manipulator has to cut the fruit or vegetable, but it starts form the beginning, that is, from the right manipulator's design choice to its correct control. Therefore principles for manipulator development were described and methods for solving inverse kinematics problem were compared and results – analyzed; main task – precise pointing at vegetable – was performed. During experiments it was stated that some methods had singularities. Tests with real vegetable showed that there are factors that affect vegetable detection.*

Keywords: robot manipulator, inverse kinematics, vegetable detection.

Introduction

It is well-known that in agricultural sector technological developments can reduce production costs (Hayashi et al., 2005). Nowadays development of robotic systems in agriculture has experienced an increased interest (Pedersen et al., 2008). In the field of robotized fruit and vegetable harvesting, the whole process can be divided into two main stages: recognition and the actual harvesting. Usually fruit and vegetable recognition is done using image processing, but harvesting - using robot manipulator.
When developing a robot manipulator, various factors must be considered, because they can have significant impact on development costs and also increase the complexity of the control system.
Motion of manipulator, like other rigid body motion when the causes of motion are not considered, is described using kinematics. Depending on what the problem is, forward or inverse kinematics is used. Forward kinematics problem is easier to solve, than inverse kinematics, because forward kinematics is based on the manipulation with the structure that is done by changes of the joint angles inside the controlled structure, but inverse kinematics is based on the direct manipulation with the end of the structure and the joint angles are derived from changes of the end of the structure (Bařinka and Berka, 2002). In case with vegetable harvesting, inverse kinematics method is used, because the target coordinates are known, but the values of robot manipulator joint angles are not. Many scientific articles provide information about several methods for solving inverse kinematics problem. These methods are such as inverse Jacobian, Jacobian transpose, analytic method and other. Different methods have been tested and compared by (Aristidou and Lasenby, 2009; Buss, 2004; Bařinka and Berka, 2002). But not all inverse kinematics methods can be used at every task.
The aim of the research is to develop robot manipulator and its motion control for vegetable harvesting. This paper focuses on robot manipulator development; analysis and comparison of inverse kinematic methods to ensure manipulator's ability to reach the target object. Target object coordinates were obtained by camera, using image processing.

Materials and methods

Manipulator's design and build

Before building manipulator, its design and materials must be chosen according to the task, this manipulator will perform. There are several manipulator design types such as polar coordinate type, articulated type, cylindrical coordinate type, Cartesian coordinate type and SCARA type (Sakai et al., 2008). Different factors characterize manipulator design. The basic mechanism of a manipulator is defined by its degrees of freedom, the type of joint, link length and offset length (Kondo and Ting, 1998).
For the task, to point manipulator's end-effector at target object, design, similar to articulated type, was chosen (Fig. 1).

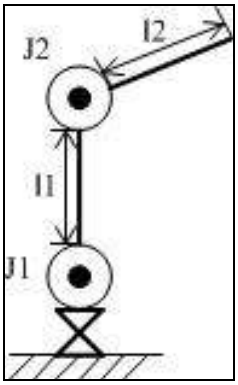


Fig. 1. **Selected manipulator design:** l1, l2 – links, J1, J2 – joints.

As it is shown in the figure, manipulator has three DOF or degrees of freedom. A degree of freedom is a joint on the arm, a place where it can bend or rotate or translate (Robot arm tutorial). All three DOFs in this case are rotational.

Different types of motors such as servo, DC, stepper, can also be used as manipulator's joints. In this research servo motors were used due to their built-in capability of moving to precise position. It is very important to choose the right motors, because if the motors are not suitable for the specific task, the whole manipulator can be damaged in the worst case scenario.

One of the main things that must be taken into consideration, when choosing a motor, is its torque. To find out the torque that is required to lift manipulator's link or "shoulder", it must be calculated (Robot statics). Torque can be calculated using formula (1):

$$M = F * l \tag{1}$$

where M – motor torque, Nm;
F – force, N;
l – distance, m.

To calculate torques, the manipulator must be stretched to its maximum length horizontally (Fig. 2), because then manipulator's motors torque will be the highest.

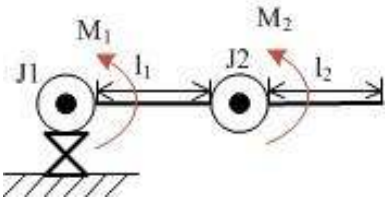


Fig. 2. **Manipulator position for motor torque calculation:** M_1, M_2 – torques, l_1, l_2 – links, J1, J2 – joints.

Torque for motor 1 is calculated in (2):

$$M_1 = \frac{l_1}{2} * G_1 + l_1 * G_m + \left(l_1 + \frac{l_2}{2}\right) * G_2 + (l_1 + l_2) * G_o \tag{2}$$

where M_1 – torque of motor 1, Nm;
l_1 – length of link 1, m;
l_2 – length of link 2, m;
G_1 – weight of link 1, N;
G_2 – weight of link 2, N;
G_m – weight of motor 2, N;
G_o – weight of object, N.

Formula above (2) can be used, if manipulator's link is homogeneous rod. If it is not, then center of mass should be determined differently (depending on the rod). And then the calculation should look like (3):

$$M_1 = a * G + l_1 * G_m + \left(l_1 + \frac{l_2}{2}\right) * G_2 + (l_1 + l_2) * G_o \tag{3}$$

where M_1 – torque of motor 1, Nm;
a – center of mass, m;
G – weight, N;
l_1 – length of link 1, m;
l_2 – length of link 2, m;
G_m – weight of motor 2, N;
G_2 – weight of link 2, N;
G_o – weight of object, N.

Torque for motor 2 is calculated in (4):

$$M_2 = \frac{l_2}{2} * G_2 + l_2 * G_o \tag{4}$$

where M_2 – torque of motor 2, Nm;
l_2 – length of link 2, m;
G_2 – weight of link 2, N;
G_o – weight of object, N.

The same thing about homogeneous rod applies to torque for motor 2. As we can see, there is weight G_o which is some object's weight. It was added for safety reasons, pretending that there is some object at the end of link 2. In this case, even though there is no object intended to be, this gives confidence about that the robot manipulator will not collapse by just lifting only these two links and one motor.

By these formulas (2-4) we can conclude that, if more DOFs are added, more complicated these formulas become, more torque is required for motors. Motor torque is also affected by link properties – for lighter weight, and shorter links smaller torque is required.

After torque calculation three motors were chosen and the manipulator was developed. Those motors are:

- two HD-6001HB,
- HITEC HS-311.

Motor physical control was ensured using microcontroller Atmega8.

Methods for solving inverse kinematics problem

Many authors divide inverse kinematics methods in some kind of groups. Methods are divided in algebraic and iterative groups by (Bařinka and Berka, 2002). Algebraic method group consist of a method which is based on basic trigonometry. The second group, iterative, consists of methods such as inverse Jacobian, cyclic coordinate descent (CCD), genetic programming, Jacobian transpose and other. These methods are called iterative, because many steps are required to solve the problem. Based on conclusions by (Bařinka and Berka, 2002), methods, such as algebraic or analytic, inverse Jacobian, Jacobian transpose, Jacobian pseudoinverse, were implemented, compared and analyzed.

Analytic method

As mentioned before, algebraic or analytic method is based on basic triangle calculations. Fig. 3 shows a situation where manipulator has reached target point.

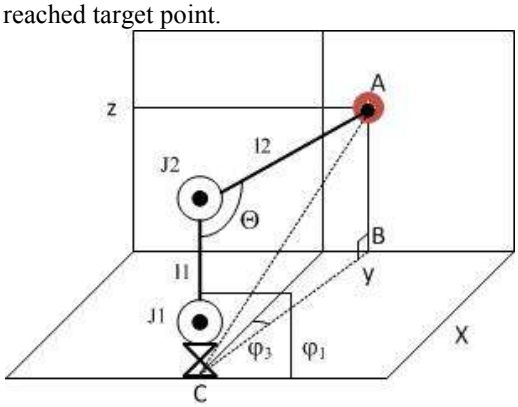


Fig. 3. **Analytic method's interpretation in 3 dimensions:** X, y, z – target object coordinates, φ_1, φ_3 – DOF angles, l1, l2 – links, J1, J2 – joints, Θ - angle between l1 and l2.

For analytic solving, first we need to create a triangle. In Fig.3 the triangle is *ABC*. When it's done, it can be transformed into two dimensional plane (Fig. 4).

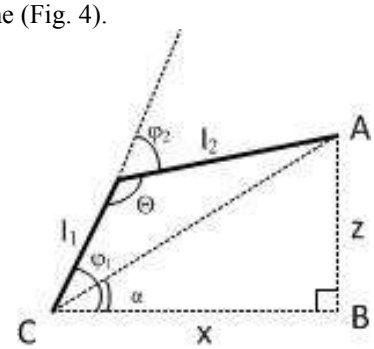


Fig. 4. **Analytic method's interpretation in 2 dimensions:** x – distance to object by *x* axis, z – distance to object by *z* axis, α – angle in triangle ABC, l_1, l_2 – links, φ_1, φ_2 – l_1 and l_2 angles respectively, Θ - angle between l_1 and l_2.

As we can see in Fig. 4, there are two triangles, where two sides are known, but the third is not. *AC* is calculated using Pythagorean Theorem. Angles such as Θ and φ_1 are calculated using the law of cosine. The equations for φ_1 and Θ are shown below:

$$\varphi_1 = \arccos\left(\frac{x}{\sqrt{x^2+z^2}}\right) + \arccos\left(\frac{x^2+z^2+l_1^2-l_2^2}{2\sqrt{x^2+z^2}*l_1}\right) \tag{5}$$

where φ_1 – link 1 angle, degrees ^{0};
x – distance to object by x axis, m;
z – distance to object by z axis, m;
l_1 – length of link 1, m;
l_2 – length of link 2, m.

$$\theta = \arccos\left(\frac{l_1^2+l_2^2-x^2-z^2}{2*l_1*l_2}\right) \tag{6}$$

where Θ – angle between l_1 and l_2, degrees ^{0};
x – distance to object by x axis, m;
z – distance to object by z axis, m;
l_1 – length of link 1, m;
l_2 – length of link 2, m.

Angle Θ is angle between links l_1 and l_2, but that is not what is needed. To calculate value of φ_2, Θ must be subtracted from 180^0 (adjacent angle feature).
The equation for solving φ_2 is shown in (7).

$$\varphi_2 = 180^0 - \theta \tag{7}$$

where φ_2 – link 2 angle, degrees ^{0};
Θ – angle between l_1 and l_2, degrees ^{0}.

To solve the third angle that represents by how much to turn the whole manipulator, equation (8) was used.

$$\varphi_3 = \mathrm{arctg}\left(\frac{(C-y)}{X}\right) \tag{8}$$

where φ_3 – manipulators turn angle, degrees ^{0};
y – object coordinate by y axis, m;
X – object coordinate by x axis, m;
C – center position by y axis, m.

It should be noted, that analytic method can be used when number of DOFs is small, for example two. In this case there were three DOFs, but this method still worked. It was because the third DOF, the base, simply turns the whole manipulator, so it was easy to solve. But when the number of DOFs is higher, these angles cannot be solved in a trivial way. Therefore more sophisticated approaches are necessary (Bařinka and Berka, 2002). Analytic method also is described by author (Yetim, 2009).
Most popular methods for inverse kinematics problem solving are those where Jacobian matrix is taking part. The Jacobian matrix $J(e,\varphi)$ shows how each component of e varies with respect to each joint angle (Rotenberg, 2005). A Jacobian matrix is nothing more than a matrix of partial derivatives of the entire chain system relative to the end-effectors (Aristidou and Lasenby, 2009). For a two dimensional robot manipulator the basic Jacobian matrix can be written as follows (9):

$$J(e,\Phi) = \begin{bmatrix} \frac{\partial e_x}{\partial \varphi_1} & \frac{\partial e_x}{\partial \varphi_2} \\ \frac{\partial e_y}{\partial \varphi_1} & \frac{\partial e_y}{\partial \varphi_2} \end{bmatrix} \tag{9}$$

where J – Jacobian matrix;
e – end-effectors' position;
Φ – joint angles.

End-effector's position is determined using equations that correspond to each end-effector's coordinate in Cartesian coordinate system. Fig. 5 shows a schematic representation how the situation (manipulator and target object emplacement) is looking like. Difference from analytic method – this method starts solving form beginning, taking into account the starting end-effector position, while in analytic method the whole process was started, assuming that manipulator had reached the object.

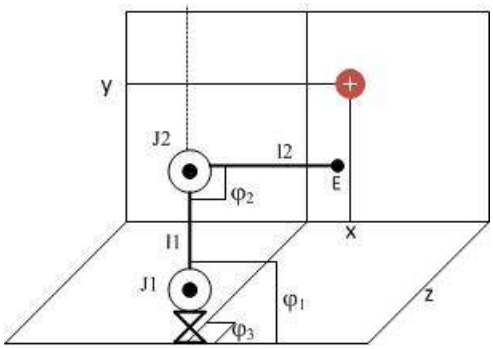


Fig. 5. **Manipulator and target object emplacement example:** φ_1, φ_2, φ_3 – DOF angles, E – end-effector, x,y,z – object coordinates, l1, l2 – links, J1, J2 – joints.

In Fig.5 we can see the starting position of manipulator. The end-effector's coordinates are calculated by equations below:

$$e_x = (l_1 cos\varphi_1 + l_2 cos(\varphi_1 + \varphi_2))cos\varphi_3 \tag{10}$$

where e_x – end-effector's coordinate by *x* axis, m;
l_1 – length of link 1, m;
l_2 – length of link 2, m;
φ_1 – link 1 angle, degrees ^{0};
φ_2 – link 2 angle, degrees ^{0};
φ_3 – manipulators turn angle, degrees ^{0}.

$$e_y = (l_1 cos\varphi_1 + l_2 cos(\varphi_1 + \varphi_2))sin\varphi_3 \tag{11}$$

where e_y – end-effector's coordinate by *y* axis, m;
l_1 – length of link 1, m;
l_2 – length of link 2, m;
φ_1 – link 1 angle, degrees ^{0};
φ_2 – link 2 angle, degrees ^{0};
φ_3 – manipulators turn angle, degrees ^{0}.

$$e_z = l_1 sin\varphi_1 + l_2 sin(\varphi_1 + \varphi_2) \tag{12}$$

where e_z – end-effector's coordinate by *z* axis, m;
l_1 – length of link 1, m;
l_2 – length of link 2, m;
φ_1 – link 1 angle, degrees ^{0};
φ_2 – link 2 angle, degrees ^{0};

The next step is to put these equations in Jacobian matrix and calculate the derivatives resulting in a matrix that can be used in selected methods.

Inverse Jacobian method

The problem, that must be solved, has a non-linear nature. Thus it cannot be solved directly, it needs to be linearized. Linearization process can be done by (13):

$$\Delta e \approx J\Delta\theta \tag{13}$$

where Δe – change in end-effector position, m;
J – Jacobian matrix;
$\Delta\Theta$ – change in joint angles, degrees ^{0}.

It tries, through *small* changes in the vector of angles $\vec{\theta}$ of the manipulator, to introduce small changes in the position of the end-effector to finally reach the desired position (Park). The change in joint angles using inverse Jacobian method, can be written like (14):

$$\Delta\theta = J^{-1}\Delta e \tag{14}$$

where $\Delta\Theta$ – change in joint angles, degrees ^{0};
J^{1} – inverse of Jacobian matrix;
Δe – change in end-effector position, m.

As described by (Aristidou and Lasenby, 2009), there are cases where Jacobian matrix cannot be inverted. This must be taken into account when trying to implement inverse Jacobian method.

Jacobian transpose method

This method is very well explained by (Aristidou and Lasenby, 2009). Basically, this method takes away problem that exists in inverse Jacobian method – possibility that matrix could not be inverted. It is so, because in

Jacobian transpose method there are no such thing as matrix inversion. Matrix transposition is performed instead of inversion. Thus this method is considered to be much faster than the inverse Jacobian method. This method is defined in (15):

$$\Delta\theta = \alpha J^T \Delta e \tag{15}$$

where $\Delta\Theta$ – change in joint angles, degrees ^{0};
α – value, usually $0 \leq \alpha \leq 1$;
J^T – transposition of Jacobian matrix;
Δe – change in end-effector position, m.

Jacobian pseudoinverse method

In pseudoinverse method change in joint angles can be expressed like (16):

$$\Delta\theta = J^+ \Delta e \tag{16}$$

where $\Delta\Theta$ – change in joint angles, degrees ^{0};
J^+ – pseudoinverse of Jacobian matrix;
Δe – change in end-effector position, m.

The solution for J^+ is described with formulas by (Yao, 2009). The main formula, which explains how to compute J^+ is shown in (17):

$$J^+ = (J^T J)^{-1} J^T \tag{17}$$

where J^+ – pseudoinverse of Jacobian matrix;
J^T – transposition of Jacobian matrix;
J – Jacobian matrix.

Authors (Aristidou and Lasenby, 2009) claim that this method has its downsides: "The pseudo-inverse method is widely discussed in the literature, however it often performs poorly because of its instability near singularities". For Inverse Jacobian, Jacobian transpose and Jacobian pseudoinverse method the algorithm is very similar. The only difference is the way, how $\Delta\Theta$ is being calculated. Activity diagram is shown in Fig. 6.

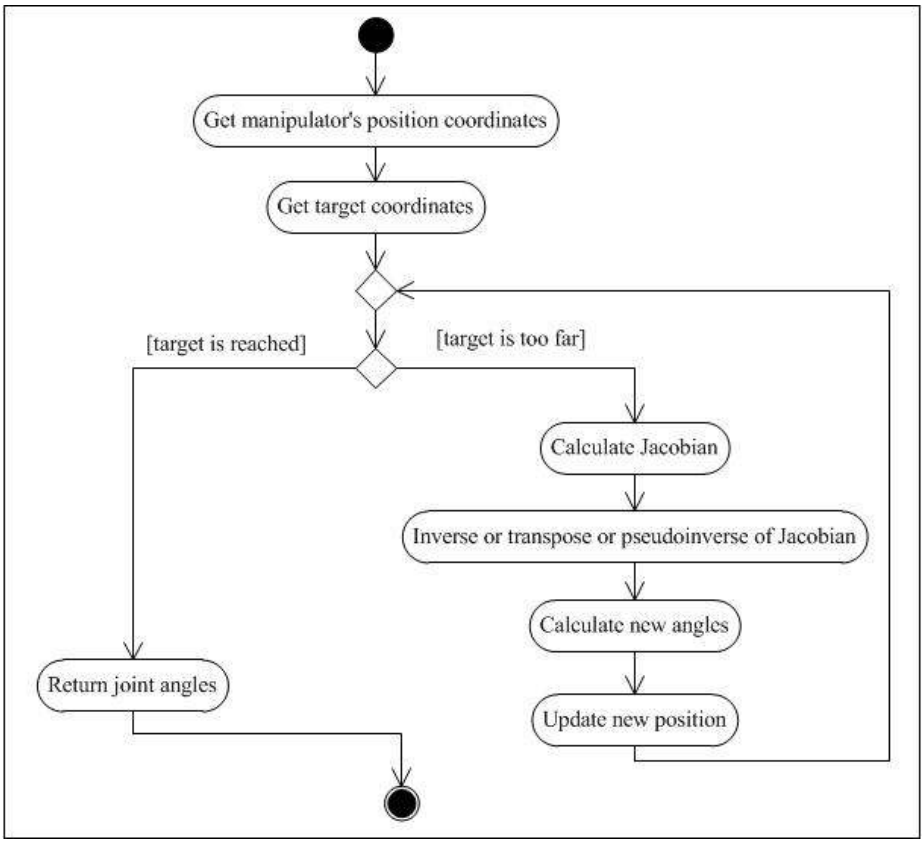


Fig. 6. **Iterative methods activity diagram.**

Usage of chosen methods

These four methods, mentioned above, were implemented into a program for vegetable recognition. The program was developed using C# programming language. For image processing and matrix calculations AForge.NET Framework was used. Images were captured using web camera with these parameters:

- sensor resolution: 1.3Mpixel,

- image sensor technology: CMOS,
- maximum video resolution: 640x480 (used in experiments: 320x240),
- maximum frame capture rate: 30fps (640x480).

Implemented methods were compared and analyzed. Results of that analysis are described in section "Results and discussion".

Results and discussion

During this research a program was developed using C# programming language. This program ensures:

- image processing,
- manipulator's angles calculations,
- sending information to microcontroller for manipulator's control.

Because of images and recognition being processed continuously and robot manipulator is controlled as soon as the angles are calculated, program was considered to be a real-time.

Comparison criteria were set for the chosen methods. It was considered, that one of the preconditions of these methods to use them in a real-time program is their execution time. If it is taking too long to calculate the joint angles, then the program will work poorly, because it must wait, while the calculation stops, to be able to turn the manipulator.

Since the manipulator is aimed to vegetable harvesting, precision is very important factor. Thus the next criterion was how precise the methods can calculate joint angles, in order to point the manipulator's end-effector to the target.

So the criteria were:

- execution time, ms;
- position error, %;

Experiments were made with C# code on 2.6GHz Athlon 64 X2.

For time calculation a time was fixed when a method starts angle calculation, and the end time was fixed when calculation was finished, so the final execution time is difference in milliseconds between these two times. This was done several times and the results were obtained with average values. The determined average execution times are shown in Table 1.

Table 1

Inverse kinematics methods average execution time

Method name	Time, ms	Iterations
Inverse Jacobian	0.59	316
Jacobian transpose	0.07	35
Jacobina pseudoinverse	7.1	507
Analytic method	0.0007	1

From Table 1 it can be concluded, that the fastest execution time had analytic method. It is so, because, this method does not belong to iterative methods group. Thus it was expected to be so. Jacobian pseudoinverse method's execution time was worst and was quite different from other.

To determine position error, a marker was attached to manipulator's end-effector. For targets red squares drawn on a paper was used (Fig. 7).

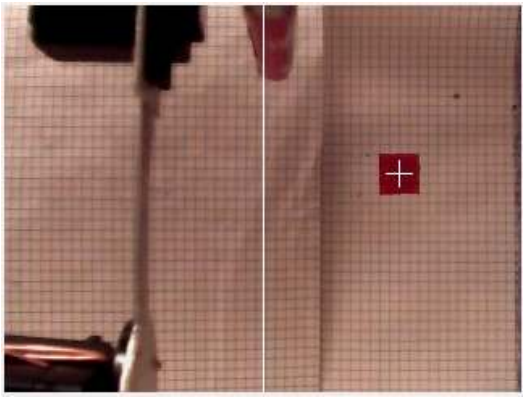


Fig. 7. **Position error detection environment.**

Error was determined by measuring the distance between manipulator's marked point and the actual target's center point, at the end total error by x and y coordinate was summed. A total of three attempts were made for each method, where in each attempt target was at different position. At the end, the average error of each method was calculated and expressed in percentages. Results are shown in Fig. 8.

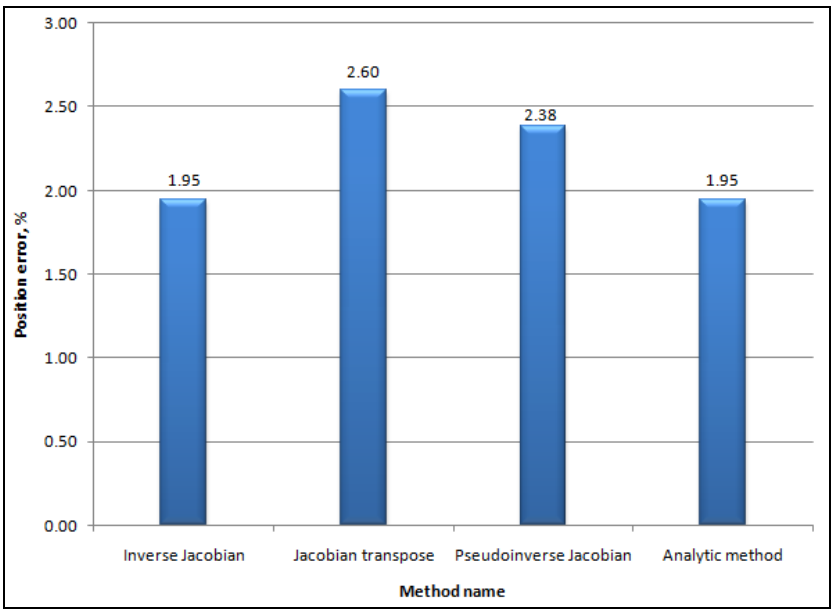


Fig. 8. **Average position error from three attempts.**

As it can be seen, the maximum precision error does not exceed 2.6%. Best results were achieved, when using inverse Jacobian and analytic method. Author (Buss, 2004) claimed, that the Jacobian transpose had the advantage of being fast, but of poor quality, the same happened here – Table 1 shows that Jacobian transpose is the fastest between iteration methods, but it has the biggest position error (Fig. 8).

During the experiments, only once a singularity, mentioned by (Aristidou and Lasenby, 2009), was detected by inverse Jacobian method, where it did not find a solution for one target position.

Referring to the paper by authors (Bařinka and Berka, 2002) about Jacobian transpose method, that some troubles like in the Jacobian inversion are still there, e. g. singularities, it must be said, that there was observed a singularity, when choosing α value – this method did not perform its task for some α values.

It is important to mention, that position error can be affected by how precise target object can be recognized and how well scaling process is done. While the experiments were performed, the error by image processing fluctuated due to light change. But these fluctuations were not very large, mostly under 0.7%. Error levels expressed in percentages at each attempt made by image processing are shown in Fig. 9.

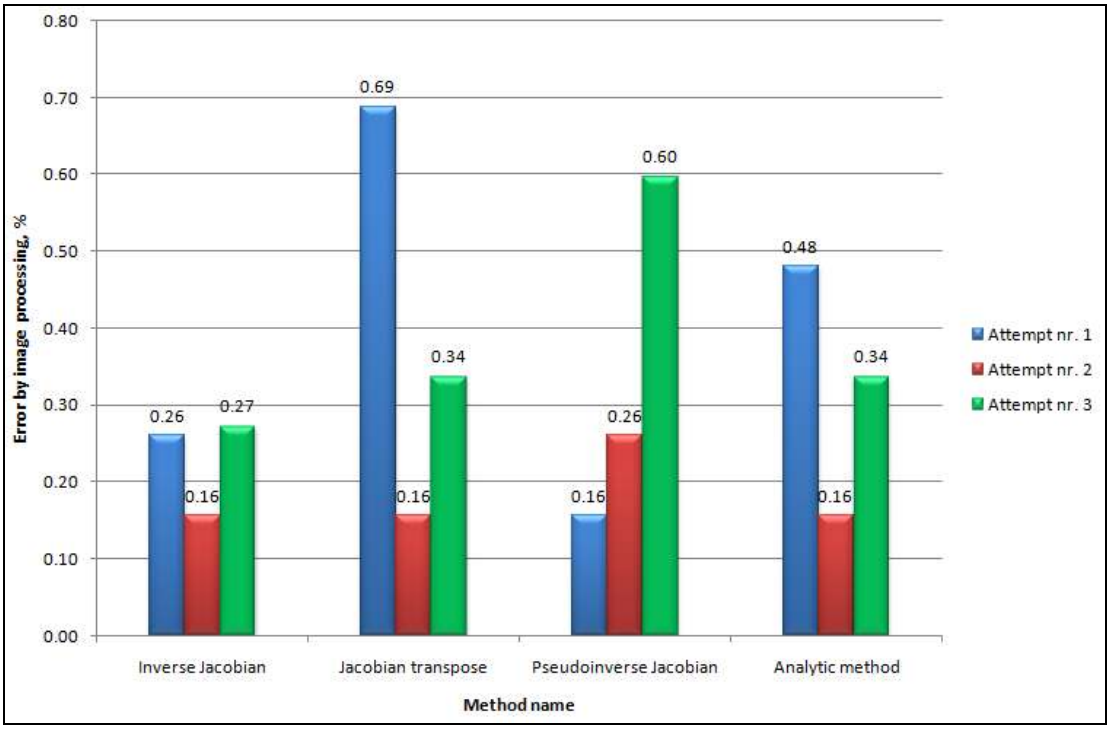


Fig. 9. **Image processing error in each attempt.**

Position error is also affected by how precise the manipulator is built.
After the analysis of inverse kinematics methods, analytic method was chosen as the suitable one and tested with real vegetable – tomato. This method was chosen because of its fast execution time and quite good precision, despite of the error made by image processing, which was higher than, when inverse Jacobian method was tested (both has the same error level though).
When the recognition and manipulator's positioning process was applied for tomato detection, some factors that affected the whole process were observed. Experiments with tomato revealed that the shape of tomato needs to take into account, because when testing the precision of inverse kinematics methods, two dimensional targets were used, but tomato is not in two dimensions. Thus during experiments often was observed situation, when manipulator hit the tomato instead of pointing to it at very close distance.
Next thing that was observed was light reflection of tomato, due to its smooth surface. In some cases this gave an impact to image processing, resulting in inaccurate coordinate detection thus affected further recognition process.
In Fig.10 tomato detection is shown. As it can be seen from side view, manipulator's end-effector does not point to tomato very precisely, reasons for that have been mentioned above.

Fig. 10. **Tomato detection:** 1. – tomato recognition (web camera view), 2. – manipulator pointing to tomato (web camera view), 3. – whole situation from side view.

Conclusion

This paper described manipulator's development principles and analysis of several methods that can be used to solve inverse kinematics problem. After inverse kinematics methods comparison and analysis, it was concluded, that the best results were obtained with analytic method, because of its very fast execution time and best error level. Analytic method's execution time was 0.0007 ms, and precision error was 1.95%. Pseudoinverse Jacobian method had the worst execution time – 7.1 ms, but Jacobian transpose method had the worst precision error – 2.6%. Analytic method was chosen to use with real vegetable recognition.
Tests with tomato showed that there are some factors that made an impact to recognition process. Thus using only image processing and methods for manipulator positioning is not enough for complete vegetable detection. To use this recognition system in a field, manipulator should be equipped with extra sensor/-s.

References

Aristidou, A. and Lasenby, J., 2009. Inverse Kinematics: a review of existing techniques and introduction of a new fast iterative solver. Cambridge University Engeneering Department. CUED/F-INFENG/TR-632. 74 p.

Bařinka, L. and Berka, R., 2002. Inverse Kinematics – Basic Methods. Dept. of Computer Science & Engineering. pp. 10

Buss, S.R., 2004. Introduction to Inverse Kinematics with Jacobian Transpose, Pseudoinverse and Damped Least Squares methods. Department of Mathematics; University of California, San Diego. 19 p.

Hayashi, S., Ota, T., Kubota, K., Ganno, K., Kondo, N., 2005. Robotic Harvesting Technology for Fruit Vegetables in Protected Horticultural Production. *FRUTIC 05,* Montpellier France. pp. 227 – 236

Kondo, N. and Ting, K.C., 1998. Robotics for Plant Production. *Artificial Intelligence Review* 12(1-3). pp. 227-243

Park, D. Inverse Kinematics. Computer Graphics, Department of Computer Science, University of Buenos Aires. 9 p.

Pedersen, S.M., Fountas, S., Blackmore, S., 2008. Agricultural Robots – Applications and Economic Perspectives. Service Robot Applications. I-Tech Education and Publishing KG, pp. 369-382

Robot arm tutorial. Available at: http://www.societyofrobots.com/robot_arm_tutorial.shtml, 04.12.2012.

Robot statics. Available at: http://www.societyofrobots.com/mechanics_statics.shtml#momentarms, 04.12.2012.

Rotenberg, S., 2005. Inverse Kinematics (part 2). Available at: http://graphics.ucsd.edu/courses/cse169_w04/CSE169_12.ppt, 04.12.2012

Sakai, S., Iida, M., Osuka, K., Umeda, M., 2008. Design and control of a heavy material handling manipulator for agricultural robots. *Autonomous Robots* 25(3). pp. 189 – 204

Yao, M., 2009. Mathematics for Inverse Kinematics. 15-464: Technical Animation. 30 p.

Yetim, C., 2009. Kinematic Analysis for Robot Arm. Yildiz Technical University, Electrical and Electronics faculty, Department of Computer Engineering. Istanbul. 34 p.

Historical Background of Scientific Activities at Faculty of Information Technologies of Latvia University of Agriculture

Aleksandrs Gailums

Department of Computer Systems, Faculty of Information Technologies, Latvia University of Agriculture, Liela str. 2, Jelgava, LV-3001, Latvia
Aleksandrs.Gailums@llu.lv

Abstract: *The aim of this article is to give an overview of development of scientific activities related to information technologies at Latvia University of Agriculture from the 1960s to nowadays. Faculty of Information Technologies was founded in 2001, but scientific activities related to information technologies began in the middle of 1960s. In 1967 the Department of the Economic Mathematical Methods and Computer Technique was established. Researches on usage of the mathematical methods developed at the new established department. The most important tasks were related to optimization of the agricultural production and cattle feeding at the collective farms. Information technologies were based on the mainframe computers. In 1972 the department stated above was renamed to Department of Economic Cybernetics, and later it was renamed to Department of Informatics. At this time several dissertations were presented. In the 1980s scientific activities were related to using personal computers. In 1992 Institute of Informatics was founded. New scientific directions were simulation modeling methods, tools and expert systems. The Faculty of Information Technologies was founded on the basis of Institute of Informatics and new scientific fields were developed.*

Keywords: Information technologies, computers, mathematical methods, scientific activities.

Introduction

By the end of the 1950s the mainframe computers were introduced in main universities and scientific institutes of republic of Latvia. In 1959 there were two computing centers founded. The first one was in Institute of Physics of Science Academy where the computer LM-3 the following year was installed and the other one was in Latvia State University where computer BESM-2 was installed (Strazdiņš, 1972).
First attempts to use Economic Mathematical Methods (EMM) and electronic computers in perspective planning of agricultural production in Latvia started in 1965 when main indices of development of agriculture for Latvia in total and for each region separately were calculated (Bite, Krastiņš, 1967). In 1967 Economic Institute of Science Academy (EISA) published the scientific issue on perspective planning of agricultural production with electronic computer BESM-2 where methodical guidelines for using EMM were directed. Experience in region Tukums served as the background for the guidelines (Lauksaimniecības ražošanas …, 1967).
That time also Latvia Academy of Agriculture (LAA) started scientific activities in this field. In the middle of 1960s there were researches related to usage of EMM and computers. In further years there were other different events and changes. In this article these events and changes are described in decades from the period of the 1960s.

1960s

In the 1960s teaching staff of several departments of LAA expressed interest in EMM and electronic computers. Moreover, updated education on computing was required for the students of all faculties of LAA. In order to satisfy these requirements it was decided to join forces of several computing enthusiasts into one department. Thereby in 1967 Department of Economic Mathematical Methods and Computer Technique (EMM and CT) at the LAA was established. Docent candidate of technical science Alberts Krastiņš was nominated to be the head of this department. The following teaching staff of LAA also joined the new department: docent Austra Brigmane, lecturer Arvīds Brūvers, lecturer Haralds Kauss. All staff mentioned above previously worked in Department of Higher Mathematics. Lecturer, candidate of agricultural science Aina Ratkeviča joined the new department from Department of Economics of the Faculty of Agriculture, assistant Zinta Ziediņa came from Department of Statistics and Accounting, lecturer candidate of agricultural science Uldis Štibe, and lecturer candidate of agricultural science Auseklis Zemītis came from Faculty of Forestry (Lauksaimniecības ekonomikas …, 1982). The new department staff focused on teaching such subjects as EMM, computing technique in engineering and economic calculations, mathematical statistics, and theory of probability.
In 1968 the Faculty of Economics of Agriculture (headed by the dean Voldemars Strīķis) on basis of several departments of the Faculty of Agronomy was founded. Department of EMM and CT was made as part of this faculty (Ekonomikas augstākā …, 2003). In this department the students were acquiring skills with the counting frames, logarithmic rules, mechanical and electro mechanical calculators, accounting machines and punch card

machines. The teaching staff of the department provided basics of computing techniques for students of all the faculties of LAA.
The department staff turned to research work: A.Krastiņš investigated methods of mathematic programming, A.Ratkeviča analyzed planning and usage of the forage, A.Brigmane – usage of statistical methods in prediction of productivity of cereals, A.Brūvers – optimization for size and location of the stock-farms.
In 1969 student Aleksandrs Gailums (author of this article) worked out diploma paper about using EMM in the planning process for forage utilization (headed by A.Ratkeviča). This mathematics task was solved by using mainframe computer BESM-4 that was placed in Latvia State University (LSU). The information of collective farm "Draudziba" in region Bauska served as the background of this task (Ratkeviča, 1970).
In 1968 the Informative Computing Center of Ministry of Agriculture (ICCA) was established (managed by director H.Kauss). The aim of this Center was to find the solutions for different tasks related to management of agriculture. The data was processed by the mainframe computer „Minsk-22". A.Gailums, A.Priedīte, V.Klešnieks after graduating of LAA started to work at Department of Optimization. The employees of this department worked under supervision of the researcher of EISA - A.Sproģis.
The farm „Zaļenieki" was the first farm where experiments of optimization planning of agricultural production were done. The experiments proved that the computer is preferred for perspective planning. New methods gave possibility to work out several solutions and to choose the best or the optimal version. The optimal solution provided that the production resources and branches are balanced. Therefore employees now were set free from exhausting work and could pay more attention to analyzes.

1970s

In the 1970s new period started. Now instead of experiments of optimization plans, the new methods were introduced permanently in many farms. The paper worked out in Economic institute and ICCA helped to manage this process (Kolhozu un …, 1971). The new process anticipated that the specialists in the farms had to fill the information tables. Those later were gathered and summarized into a standard matrix in the computer. The researchers A.Gailums, Ē.Indāns, V.Klešnieks, A.Priedīte did methodical work for this process. In the 1970s several software packages of planning and accounting were used in the collective farms. The most important packages were dedicated to optimization of the agricultural production, optimization of the cattle feeding, farm accounting, herd work, traumatism and optimization of use of fertilizers (Kipere, 1971).
In 1972 EMM and CT Department was renamed to Department of Economic Cybernetics. Docent A.Ratkeviča (1972-77, 1984-86) and docent A.Brūvers (1978-83) were managing this department.
The computer „Minsk-22" provided vast possibilities in research work for the staff and the students of LAA. For example, A.Ratkeviča investigated optimization of forage stocking farms in the winter period and optimization of food ration for cattle (Ratkeviča, 1970) A.Brigmane started to use the statistic methods (correlation analyses, covariation analyses) for forecasting the productivity of cereals (Brigmane A., Gūtmane B., 1975). She also wrote an article „Biometric methods in selection" where she spoke about researches on standard deviations, variations, coefficient of correlation and analyses of dispersion (Lindermanis, Brigmane, 1970).
The researches of A.Brūvers were related to the expert estimate methods used for forecasting the efficiency of agricultural production (Brūvers, 1977). He used the computer Nairi-S for his data processing. U.Štibe investigated optimization for using the mineral fertilizers in the farms (Štibe, 1973). A.Krastiņš turned to problems of the mathematic programming – teaching it in the agricultural high schools (Krastiņš, 1975). A. Brūvers (Brūvers, 1970) and A.Brigmane (Brigmane, 1972) defended their thesis of candidate of economics science.
In 1974 the 29th Scientific Conference of LAA took place. A. Ratkeviča chaired the section of Economic Cybernetics. The researchers presented the researches of this department. A.Gailums and A.Sprogis (EISA) gave report about the development of automatic planning system in agriculture of Latvia (LLA 29. zinātniski …, 1974).
In 1975 A.Gailums was transferred from ICCA to the Department of Economic Cybernetics to work as an assistant. A.Gailums delivered lectures and practical works in course „System of automatic data processing" and „Computer technique in engineering-economic calculations". His researches were related to the optimization of perspective planning in the farms.
In 1976 two textbooks were published: A.Ratkeviča „Mathematic modeling of agriculture" (Ratkeviča, 1976) and A.Krastiņš „Mathematic programming" (Krastiņš, 1976). In 1979 A. Brūvers worked out teaching material „Basic of programming for computer Nairi-S" (Brūvers, 1979).
In the 1970s several automatic management systems related to the agriculture were implemented. For example, information system for herd-work of cattle „Selex" was launched (Arhipovs, 1979). The group headed by prof., Doctor of Economics B.Treijs worked on using the linear programming in the agricultural perspective planning.

1980s

In the beginning of 1980s the mainframe computers were replaced by personal computers. Therefore all the tasks had to be adjusted to the personal computers specifics.

The automated working places were created in the collective farms. For example, in the mid-1980s the Robotron 1720 in the collective farm "Taurene" of the district Cēsis performed several accounting tasks: cattle breeding accounting and electrical resources accounting, storehouse accounting (Strauts, 1987).
Same time, also Department of Economic Cybernetics installed the first personal computers. The first ones were Robotron-1715, Iskra-1817 and Pravec, later followed by IBM PC. Also the minicomputers such as Nairi-3, Nairi-S, Iskra-226 and the programming calculators were used. The laboratory of the Department of Accounting used the minicomputer M-5010. Mostly the programming language BASIC was used. All teaching staff took part in on-line learning courses of BASIC. The students mastered different Office applications: word processor, spreadsheet and databases. The subject "Informatics" was delivered for the students in all the faculties. The optimization of food rations started to solve by help of the computers Pravec, Iskra-226 and CM-4. The methodic was made and introduced by the researcher A.Ivane.
The Department staff continued the scientific researches. A.Brigmane researched how to improve the function of grain production (Treijs, Brigmane, 1980) and A.Brūvers researched analyze of the factors of the experts' estimations. (Brūvers, 1980)
A.Gailums defended his doctor thesis that was supervised by professor B.Treijs (Gailums, 1981). The economic mathematical model was solved on the computer „Siemens 4004" at State Planning Institute in the city of Rīga. The data forecast for model was solved on Nairi-S. A.Ratkeviča defended her Doctor habil. thesis of economics science on subject "Planning of forage with electronic computers" (Ratkeviča, 1989).
In 1987 the Faculty of Agricultural Economics moved to a new place out of the main LAA building. But the Department of Economics Cybernetic stayed. It was now subordinated to the Rector of LAA. The head of department that time was docent J.Beidermanis. Same time ICCA moved to the Ministry of Agriculture in city Rīga. Thereby Department of Economic Cybernetics occupied these rooms in basement of the palace.

1990s

In 1990 the Department of Economic Cybernetic was renamed to Department of Informatics. The head of the Department was prof., Dr.habil.sc.ing. Pēteris Rivža who managed the Department of Mathematics beforehand
In 1992 the Institute of Informatics was founded on basis of the Department of Informatics. The new Institute included the Department of Informatics, the Department of Mathematics and three divisions. The first one was the Division of Information Systems headed by V.Birkants (1992-93), A.Ivane (1993-98), S.Sproģe (since 1998). The second was the Division of Computer Network Service headed by A.Paura and the third was the Centre of Information Technologies headed by Ģ.Kazainis. Dr.habil.sc.ing. P.Rivža was the director of the Institute of Informatics (Rivža …, 1999).
The Department of Informatics worked on such subjects as informatics, theory of probability and mathematical statistics, quantitative analysis methods, econometric, programming language HTML, Web pages, control systems of databases and communication technologies. New internet classroom was founded at the Institute. The teaching staff continued to perform the methodical and scientific work.
The aspects of methodology and mathematical modeling of grain drying and storage processes have resulted in two doctoral theses (Arhipova, 1994; Āboltiņš, 1993) and one doctoral habil.thesis (Rivža, 1995). In 1999 L.Paura defended doctoral thesis about development of model of estimating of pedigree bulls (Paura, 1999). L.Ramute researched using of cluster analysis of grouping rural municipalities (Ramute, 1996).
Several methodical materials were issued: "Micro calculators programs of counting for agriculture" (Rivža, 1993), programming micro computer "Electronic B3-34" (Ziediņa 1993) and "Working with word processor MS Word 6.0" (Gailums, Dmitrijeva, 1996) and "Working with spreadsheet MS Excel" (Gailums, Dmitrijeva, 1999). In the 1990s the cooperation with Estonian, Lithuanian Swedish, and Netherland and Italic researchers was developed.
The 1990s are characterized by the forming of peasant farms, which were founded in Latvia as a result of the Land reform. The personal computers were a powerful tool for the data processing and the problem solving. The farmer as a computer user was becoming more and more directly involved in the process of information through the personal computers.
In January, 1991 Latvia Agricultural Advisory and Training Center was established by the Ministry of Agriculture. Its purpose was to provide the trainings and consultations for the farmers and rural enterprises. It also offered such software packages as optimization of food ration, optimization of fertilizers and accounting for peasant farms.

2000s

In the study year 2000 / 2001 the Institute of Informatics set up the academic bachelor study program "Computer Control and Computer Science". The dean of the Faculty of Automatic and Computer Technique of Riga Technical University prof. Jānis Grundspeņķis helped to form this new study program.
In the study year 2001 / 2002 the second study program – professional bachelor study program "Programming" was launched. The vice president of the company "Exigen Services Latvia" assoc. prof. U.Smilts was consulting and helping in the forming process of this study program.

The new launched programs initiated the foundation of the Faculty of Information Technologies (FIT). This Faculty was founded in 2001 on bases of the Institute of Informatics. The deans of the Faculty were prof. P.Rivža (2001-02), prof. I.Arhipova (2002-08), prof. U.Iljins (since 2008).

The Department of Informatics was divided into two departments: the Department of Computer Systems headed by assoc. prof. Dr.oec. A.Gailums (2001-06), assoc. prof. Dr.sc.comp. R.Čevere (since 2006) and the Department of Control Systems headed by assoc. prof. Dr.sc.agr. L.Paura. The FIT also included the Department of Mathematics headed by prof. Dr.sc.ing. A.Āboltiņš (1990-2007), prof. Dr.sc.paed. A.Zeidmane (since 2008) and also included the Department of Physics headed by prof., Dr.habil.sc.ing. U.Iļjins (1994-2008), assoc. prof. Dr.sc.ing. U.Gross (since 2008).

In 2005 the Master study program "Information Technologies" was developed. A significant moment in the scientific activities of FIT was in 2006 when the first graduates of the FIT master study program started their studies in the newly developed doctoral study program "Information Technologies".

The FIT was organizing different international scientific conferences. The first one (2004) and the second one (2006) took place under the title "Information and communication technologies for rural development". Since 2008 the conferences were titled "Applied information and communication technologies".

Within the framework of the collaboration between the Forest Faculty and the FIT, an interdisciplinary team of researchers from the scientific disciplines of information technologies and forestry was established. A co-operation was established with the Faculty of Rural Engineering into researching the qualities of a foam plaster and other molded building materials. Also co-operation with the Institute of Microbiology and Biotechnology of the University of Latvia was established (Stalidzāns, Arhipova, 2009). The project "ICT-AGRI", which has been implemented within the 7th framework program in the period of 2009-2013, provides an opportunity to integrate into the relatively specific sphere of the agricultural application of IT at an international level.

The Department of Mathematics has made close cooperation with the Faculty of Mathematics of the Estonian University of Life Sciences. There is also a collaboration agreement for science work with the Department of Physics of the Lithuania University of Agriculture. The collaboration has been established with such employers as SIA "Exigen Services Latvia", SIA "Lattelecom Technology" and SIA "Microsoft Latvia".

The four departments of the FIT provide vast variety of scientific subjects. The main activities of the Departments are described below in the article.

The main directions of scientific activities at the Department of Computer are: computer control systems (leading researcher – E.Stalidzāns): systems and synthetic biology (leading researcher – E.Stalidzāns); development of information and communication technologies in Latvia (leading researcher – P.Rivža); program engineering (leading researcher – R.Čevere); agricultural information systems (leading researcher – A.Gailums); modeling of planning and managing forestry (leading researcher – I.Šmits); model based precision computer control of the multiobject biosystem (leading researcher – A.Zacepins), Modelling of autonomus hybrid power supply control systems (leading researcher - V.Osadčuks), cross-cultural Web information systems design (leading researcher - G.Vītols).

The main directions of scientific activities at Department of Control Systems are:

applications of information technologies in forestry (leading researcher – I.Arhipova); bioinformatics (leading researcher – L.Paura); methods of statistics and region analyses (leading researcher - L. Ramute); modeling of waste pollution (leading researcher - L.Bērziņa).

The main directions of scientific activities at Department of Mathematics are: pedagogic (leading researcher – A.Zeidmane); modern elementary mathematics and didactics of mathematics (leading researcher – L.Ramāne); educational management (leading researcher – A.Vintere). Dr.silv. professor emeritus R.Ozoliņš has been carrying out research into the forestry science already since 1970. In 2007 the Ministry of Agriculture of the Republic of Latvia awarded R.Ozoliņš with the highest award of the forest industry "Gold Cone" for his lifetime contribution to the development of the forest science in Latvia.

The main directions of scientific activities at Department of Physics are: heat and mass transfer (leading researcher – U.Iljins); research on solar collectors (leading researcher – U.Gross); influence of physical parameters of atmosphere on use of solar energy in several collectors (leading researcher – I.Pelēce).

Since the foundation of the FIT there are 12 staff members who have defended their doctoral thesis: (Gross, 2002), (Ramāna, 2004), (Kopeika, (2007), (Ramute, 2008), (Sergejeva, 2010), (Atslēga, (2011), (Pelēce, 2011), (Vronska, 2011), (Osadčuks, 2012), (Gedrovica, 2012), (Mozga, 2012), (Vītols, 2012).

In 2010 the Promotion Council of Information Technologies was founded and it consists of LUA assoc. prof. Dr.sc.comp. R.Čevere (head of the council), LUA prof. Dr.sc.ing. I.Arhipova, LUA assoc. prof. Dr.sc.ing. E.Stalidzāns, LUA assoc. prof. Dr.oec. A.Gailums, LU prof. Dr.habil.sc.comp. J.Borzovs, RTU prof. Dr.habil.sc.ing. J.Grundspeņķis, LU assoc.prof. Dr.sc.comp. J.Vīksna, and the scientific secretary of the council - T.Tabunova. The first promotion theses were defended in 2012 – I.Mozga (Mozga, 2012) and A.Cīrulis (Cīrulis, 2012) for acquiring Doctoral Degree in the field of information technology.

Conclusions

The development of the scientific activities at the FIT was closely linked to the development of computing techniques and to the changes in the rural areas.
The Department of EMM that was founded in 1967 and involved 7 teachers over the decades developed into the FIT where a lot of employees have made their career. There were several important steps in this development: first there was the Department of EEM and Computing Techniques, later – Department of Economic Cybernetics, then – Department of Informatics, Institute of Informatics and Faculty of Information Technologies. In the field of science, the ITF is successfully cooperating with other faculties of the LUA, other higher educational institutions of Latvia and also with partners in other European countries.

References

Arhipova, I., 1994. Graudu tīrāmo mašīnu sietu izvēles metodika, izmantojot ekspertu sistēmu tehnoloģiju. Disertācijas autoreferāts doktora inženierzinātņu grāda iegūšanai. Jelgava, LLU, 116 lpp.

Arhipovs, P., 1979. Automatizētā vadīšanas sistēma lauksaimniecībā. Rīga, Liesma, 157 lpp.

Atslēga, S., 2011. Bifurcations in nonlinear boundary value problems and multiplicity of solutions. Summary of the Doctoral thesis. Daugavpils, p.84.

Āboltiņš, A., 1993. Bieza, nekustīga graudu slāņa kaltēšanas procesa ar aktīvo ventilēšanu risināšanas metodika. Disertācijas autoreferāts doktora inženierzinātņu grāda iegūšanai. Jelgava, LLU, 114 lpp.

Bite ,V. and Krastiņš, O., 1967. Lineārā programmēšana lauksaimniecības plānošanā. Rīga, Liesma, 142 lpp.

Brūvers, A., 1970. Liellopu fermu izvietojuma plānošanas ekonomiski matemātiskās metodes. Disertācijas autoreferāts ekonomikas zin.kand. grāda iegūšanai. Odesa, 25 lpp.

Brūvers, A., 1977. Ekspertu vērtējuma metode un tās lietošana lauksaimniecīskās ražošanas rādītāju prognozēšanā. LLA raksti, 94.sējums, 53-60 lpp.

Brūvers, A., 1979. Programmēšanas pamati elektronu skaitļošanas mašīnai Nairi-S. LLA, 41 lpp.

Brūvers, A., 1980. Faktoranalīzes pielietošana ekspertu vērtējumos. LLU raksti, 184.sējums, 42-46 lpp.

Brigmane, A., 1972. Augkopības ražošanas funkciju pētīšana, pielietojot kovariācijas analīzes metodi.

Brigmane, A. and Gūtmane, B., 1975. Pētījumi par zemes kvalitātes un ekonomiskā novērtējuma sakarībām. LLA raksti, 76.sējums, 47-52. lpp.

Cīrulis, A., 2012. Virtuālās realitātes izmantošana tehnoloģisko procesu vadības apmācības pilnveidošanā. Promocijas darba kopsavilkums Dr.sc.ing. zinātniskā grāda iegūšanai. Jelgava, 111 lpp.

Ekonomikas augstākā izglītība – laukiem. 2003. LLU, Ekonomikas fakultāte, Jelgava, 331 lpp.

Gailums, A. and Dmitrijeva, T., 1996. Darbs ar teksta redaktoru MS Word 6.0, LLA, Jelgava, 139 lpp.

Gailums, A. and Dmitrijeva, T., 1999. Elektroniskā tabula MS Excel iesācējiem. LLA, Jelgava, 56 lpp.

Gailums, A., 1981. Lauksaimniecības ražošanas tehniskā nodrošinājuma sistēmas ilgtermiņa plānošanas metožu pilnveidošana. Disertācijas autoreferāts ekonomisko zinātņu kandidāta grāda iegūšanai. Jelgava, LLA, 1981, 22 lpp.

Gedrovica, I., 2012. Pastry products enriched with dried Jerusalem Artichoke (Helianthus Tuberosus L.) powder. Summary of the Doctoral thesis. Jelgava, p.42.

Gross, U., 2002. The flash photolysis method in the lodine – ozone mixture photochemical research. Summary of the Doctoral thesis. Jelgava, p. 36.

Kipere, Z., 1971. Cilvēki un mašīna, laikraksts Padomju Jaunatne, 1971.gada 24.oktobris.

Kolhozu un padomju saimniecību perspektīvo plānu galveno rādītāju optimizācija. 1971. Latvijas PSR ZA Ekonomikas institūts, Lauksaimniecības ministrija, Informatīvais skaitļošanas centrs, Rīga, Zvaigzne, 270 lpp.

Kopeika, E., 2007. Information and telecommunication technologies for the regional development of Latvia. Summary of the Doctoral thesis. Jelgava, p. 48.

Krastiņš, A., 1975. Matemātiskās programmēšanas mācīšana lauksaimniecības augstskolās. LLA raksti, 120.sējums, 53-57. lpp.

Krastiņš, A., 1976. Matemātiskā programmēšana. Rīga, Zvaigzne, 196 lpp.

Lauksaimniecības ekonomikas fakultāte. 1982. Doc. Strīķa redakcijā, Rīga, Avots, 102 lpp.

Lauksaimniecības ražošanas perspektīvā plānošana. 1967. Latvijas PSR Zinātņu Akadēmijas Ekonomikas institūts. Rīga, Zvaigzne, 270 lpp.

Lindermanis, J., Brigmane, A., Kalniņš, L., Holms, I., Rubenis, J., 1970. Vispārējā laukaugu selekcija. Rīga, Zvaigzne, 362 lpp.

LLA 29. zinātniski praktiskā conference. 1974. LLA, Jelgava, 41.lpp.

Mozga, I., 2012. Steady state optimization procedure of biochemical networks. Summary of the Doctoral thesis, Jelgava, p. 57.

Osadčuks, V., 2012. Modelling of autonomus hybrid power supply control systems. Summary of Doctoral thesis, Jelgava, p. 166.

Paura, L., 1999. Latvijas apstākļiem piemērota buļļu ciltsvērtēšanas modeļa izstrādāšana. Summary of doctoral thesis. Jelgava, LLU, pp. 46.

Pelēce, I., 2011. Influence of physical parameters of atmosphere on use of solar energy in several collectors. Summary of the Doctoral thesis. Jelgava, p. 35.

Ramāna, L., 2004. Method of invariants in elementary mathematics and its role in improvement of high-school mathematics curricula. Summary of the Doctoral thesis. Rīga, p. 52.

Ramute, L., 2008. Analyse and optimization of administrative teritorial dividing of Latvia. Summary of the Doctoral thesis. Jelgava, p. 63.

Ramute, L., 1996. Cluster analyses for the classification of rural communities. LLU raksti, 284.sējums, 99-102 lpp.

Ratkeviča, A., 1989. Lopbarības izlietojuma plānošana lauksaimniecības uzņēmumos Latvijas PSR ar ESM palīdzību. Disertācijas autoreferāts ekonomisko zinātņu doktora grāda iegūšanai. Jelgava, 71 lpp.

Ratkeviča, A., 1976. Ražošanas matemātiskā modelēšana lauksaimniecībā. Rīga, Zvaigzne, 219 lpp.

Ratkeviča, A., 1970. Skaitļošanas tehnikas pielietošanas saimnieciskajam aprēķinam metodiski jautājumi. LLA raksti, 28.sējums, 121-126. lpp.

Rivža, P., 1995. Graudu pirmsapstrādes sistēmas analīzes, modelēšanas un optimizācijas metodika. Disrtācija habilitētā doktora inženierzinātņu grāda iegūšanai. Jelgava, LLU, 329 lpp.

Rivža, P., 1999. Informātikas institūts. Lauksaimniecības augstākā izglītība Latvijā, LLU, Jelgava, 91-92 lpp.

Rivža, P., 1993. Lauksaimnieku aprēķinu programmes mikrokalkulatoriem. LLU, Informātikas katedra, Jelgava, 86 lpp.

Sergejeva, N., 2010. The spectral properties of nonlinear BVP for the second order ordinary differential equations. Summary of the Doctoral thesis. Daugavpils, p. 32.

Stalidzāns, E., .Arhipova, I, .Iljins, U, Paura, L., Vintere, A., Zeidmane, A., Zujevs, A., Ozoliņš, R., Rivža, P. 2009. Information Technologies step into New Scientific areas. Latvia University of Agriculture – 70. Proceedings of International Scientific Conference. Jelgava, pp. 89-101.

Strauts, A., 1987. Elektronskaitļotājs kolhozā. Lauku dzīve, 1987.gada maijs, 15.lpp.

Strazdiņš, I., 1972. Elektroniskās skaitļošanas mašīnas. Rīga, Zvaigzne, 340 lpp.

Štibe, U., 1973. Minerālmēslu izmantošanas optimālo plānu sastādīšana saimniecībām. LLA raksti, 66.sējums, 73-78. lpp.

Treijs, B., Brigmane A., 1980. Graudu ražošanas funkcijas pilnveidošanas iespējas. LLU raksti, 184.sējums, 51-57 lpp.

Vītols, G., 2012. Cross – cultural Web information system design. Summary of the Doctoral thesis. Jelgava, p. 63.

Vronska, V., 2011. Development of perspective teacher ICT integration skills household and home economics education. Summary of the Doctoral thesis. Jelgava, p. 74.

Ziediņa, Z., 1986. Programmējamais mikroskaitļotājs “Elektonika B3-34”. LLA, Jelgava, 25 lpp.

New wireless sensor network technology for precision agriculture

Maris Alberts[1], Ugis Grinbergs[1], Dzidra Kreismane[2], Andris Kalejs[1], Andris Dzerve[1], Vilnis Jekabsons[1], Normunds Veselis[1], Viktors Zotovs[1], Liga Brikmane[1], Baiba Tikuma[2]

[1]*Institute of Mathematics and Computer Science of the University of Latvia, Riga, Raiņa blvd. 29, LV-1459, Latvia*
[2]*Latvia University of Agriculture , Jelgava, Liela str. 2, LV3001, Latvia*
ugis.grinbergs@lumii.lv, dzidra.kreismane@llu.lv

Abstract: *Institute of Mathematics and Computer Science (IMCS) of the University of Latvia is implementing the project "Development of Long Range Wireless sensor network for precision farming applications in Latvia" (The Project). The Project has two main directions of research. The first is development of new long range Wireless Sensor Network (WSN) nodes providing radio link in a long distances (more than 300m), the second is development of energy efficient operating system (FarmOS) for large scale WSNs with main focus on robust easy to use agricultural applications. Development of current WSN nodes is based on Texas Instrument's hardware and AgroSeNET technology drafted by Cominfo Inc. FarmOS is being developed in cooperation with Institute of Electronics and Computer Science. In framework of the Project IMCS is implementing field trial of large scale Long Range WSN technology for automated radiation frost protection of cranberry fields. FarmOS supports mash topology of WSN. For small networks star and multihop topologies could be used. The main goal of the Project is to build WSN prototype with newly designed 50 long range nodes running FarmOS to provide automated cranberry field protection with intelligent radiation frost prediction and decision making features. Depending on data gathered in the real time the intelligence of the system will provide immediate decision whether fog generator, water spryer system or just wind blowers would be chosen. Farmer's end user client software will deliver Software as a Service concept.*

Keywords: Wireless sensor networks, Precision Farming, OGC, INSPIRE, Environmental monitoring.

Introduction

Global climate changes. Fossil fuel, extensive land use, and agriculture are three main causes of the increase of greenhouse gases observed over the past 250 years. Intergovernmental Panel on Climate Change (IPCC) agreed that about 25% of carbon dioxide emissions are produced by agricultural sources, deforestation, use of fossil fuel-based fertilizers, and burning of biomass. Conventional tillage and fertilizer application release about 70% of the nitrous oxide emissions. On one hand over the past centuries substantial increase in crop yield is achieved with development of technologies on the other hand intensive agricultural methods have detrimental effects on the environment (Agriculture, 2012).

On one hand intensive soil tillage reduces soil organic matter through aerobic mineralization, on the other hand low tillage and the maintenance of a permanent soil cover increases soil organic matter. No or low tilled soil conserves the structure of soil for fauna and related macrospores to serve as drainage channels for excess water. Surface mulch cover protects soil from excess temperatures and evaporation losses and can reduce crop water requirements by 30 percent. Conservation and organic agriculture that combine zero or low tillage and permanent soil cover are facilitating adaptation options for soil ability to increase organic carbon level reduce mineral fertilizers use and reduce on farm energy costs (Adaptation to climate..., 2007). Autonomous adaptation is the farmer reaction to precipitation patterns changing crops or uses different harvest and planting or sowing dates. Long term adaptation includes structural changes in land use to increase yield under new conditions, to bring into effect new technologies and land management and water use efficiency related techniques.

Early warning and risk management systems (EWIS) are efficient contributors that can further adaptation to climate change. Important information on databases (DB) are a historical climate data archives, an archive on climate impacts on agriculture, monitoring tools using systematic meteorological observations, climate data analysis, information on the characteristics of system vulnerability and adaptation effectiveness such as resilience, critical thresholds and coping mechanisms. Food and Agriculture Organization of the United Nations (FAO) is a leader in implementation new data format standards of new data types and specific tools (methods and software) and methods such as data interpolation in time and area, and analysis tools at different levels. These activities need to focus on securing agricultural productivity in a sustainable way. Familiar and well-considered are EWIS and Disaster Information Management System (DIMS) that can be used for estimation while contributing to disaster readiness and elimination of potential risks.

Local climate changes. In the Climate Change Mitigation Policy for Latvia among other goals to achieve is increase of efficient and rational use of resources in agriculture and implementation activities such as promotion of environmentally sound agricultural methods that reduce direct greenhouse gas emissions. The main aim of

agriculture is sustainable use of agricultural resources by development of environmentally friendly agriculture and promotion of Good agricultural practices (Agriculture, 2012).

Farmers of Latvia climate changes perceive as uneven rainfalls during vegetation period, longer periods of autumn, a very early or late spring, sustained rain or drought and winter thaws. This creates the need for more flexible choice of crop varieties and the application of effective production technologies. The main adaptation measure is to vary inputs of agrochemicals and amounts of fertilizer applications, alter time of application and vary amount of chemical control of pests.

Global society changes. IPCC is the leading international body for the assessment of climate change. IPCC was established by the United Nations Environment Program and the World Meteorological Organization to provide the world with a comprehensive assessment of the current situation of climate change and its potential environmental and socioeconomic impacts (Managing the Risks...., 2012).

In next few decades growth of middle class people in the world is expected from 1 900 million in year 2012 to 4 800 million in year 2030. To feed the worlds whole population in the year 2030, predicted to be 8 500 million, current food production will need to double in order to meet minimum requirements. Huge increase in global food demand projected for the next 20 - 30 years poses immense challenges for the sustainability both of food production and terrestrial and aquatic ecosystems and the services they provide to society. Farmers as main managers of global cultivated lands could change, eventually irreversibly, the surface of the planet in the next decades.

Agriculture today feeds 6 000 million people. Global grain crop production has doubled in the past 40 years mainly from increased yields resulting from greater inputs of fertilizer, water and pesticides, new crop strains, and other technologies of the „Green Revolution". Due to that prior task for agricultural producers is food security and product quality (Tilman et al., 2002).

Considering large area of agricultural lands new technologies are demanded for collecting sensitive data and evaluation of these data.

ICT adoption in rural community. As it is said before the only way to raise productiveness of agriculture is adoption of new ICT technologies in rural community. Adoption of ICT enabled information systems for agricultural development and rural viability is a strategic concern worldwide. Furthermore adoption of ICT in rural community is the only way to avoid irreversible and adverse effects on the Earths nature.

ICT for precision farming. The application of GNSS sensors and machine control to agriculture, with a prescriptive approach that matches maps of the field nutrients and soil condition, has had a valuable impact on increased yields and cost savings for farmers.

An essence of precision agriculture solutions is site specific crop management. That incorporates different aspects, such as monitoring soil, crop, and climate in a parcel and extrapolating the results to complete field. Precision farming provides decision support systems (DSS) with actual data for taking differential actions, for instance, data necessary for variable rate application of fertilizer, lime, and pesticide, or for tillage, or sowing rate in the real time.

WSN for precision farming. As a rule industrial agriculture operates with pretty large fields from tenth till hundreds of square kilometers. It makes cost rise skyrocket for any technology used. Large scale low-cost WSNs obviously are the only way for large field condition monitoring in real time thus they become a part of future farm as the key component for future Internet of Things. This article deals with design and implementation of future Things of Internet.

Materials and methods of research

As follows from the said above development of technologies for large scale networks becomes crucial for sustainability of today's agriculture.

During the current research specifications of hardware, WSN operating system, and supplementary software specifications were assessed and first drafts of them created. In terms of research sensors, technologies, microcontrollers, and available WSN radios part base selected. Together with farmer usefulness of gathered data for particular applications evaluated.

Fig. 1 shows organization of the experiment. A mock up of WSN prototype consisted of 6 sensor boards and 1 gateway, namely AgroSeNET.

The equipment was installed in a cranberry parcel for monitoring processes of **radiation frosts**. The nodes were manually localized and positioned using A-GPS and placed on map of the field. The AgroSeNET nodes were equipped with dual sensor for registration both temperature and relative humidity of air, and sensor for registration temperature of cranberry leaves. In addition some nodes were equipped with sensors registering luminosity, air pressure, precipitation, wind strength and direction of wind.

Previous researches show that radio coverage of nodes is significantly reduced when crop is flourishing (Goense et al., 2005). To avoid impact of crop canopy of leaves on communications the radios of nodes were installed at height of 200cm while the sensors were installed at height of 70cm and in the level of cranberry leaves. Distance between nodes was about 200m. A number of sensors for soil humidity measurement were deployed in the field also.

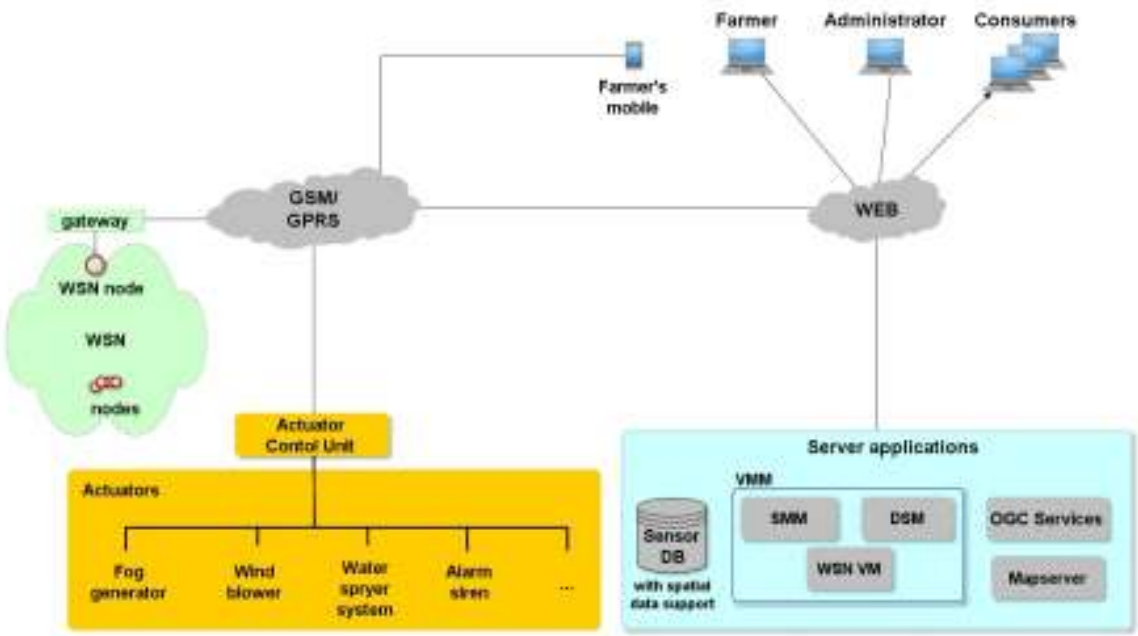


Fig. 1. **Block schematics of the experiment**.

WSNs have emerged as a result of recent advances in low-power digital and analogue circuitry, low-power RF design and sensor technology. Sensor networks are distinct from traditional computing domains. WSNs are composed of motes that are standalone pieces of hardware equipped with a tiny computer power and a radio transmitter. One of the greatest challenges that these types of networks have to face is the energy consumption requirements. Since the network nodes are located in the open and have no accessible energy sources they have to rely on limited size batteries or have they own energy scavenger ability from solar panels or wind turbines. Nevertheless WSN nodes must be built as energy saving as possible. This has led to developments of communication protocols that balance transfer speed with battery consumption like the IEEE 802.15.4, the ZigBee or the 6loWPAN. Other network protocols used are the MyriaNed or the DASH7. The protocols mentioned can provide network with communication ability. However these protocols are not operating systems and they are not enough functionality for use for large scale WSNs.

Sensors and their interfaces. There is a wide range of sensors used for agriculture. Sensors could be used for measuring atmospheric data, soil characteristics, yield qualities, and several plant attributes. Among sensors used for capturing yield qualities we can enumerate several technologies used for measuring agricultural inputs and outputs such as: impact, comb, optical, radiometric, load cell, and torsion/deformation sensors. Besides them we can place activity sensors in the harvester's head, humidity sensors, GPS positioning (DGPS differential GPS or RTK real time kinetics), tilt sensors, quality sensors.

Atmospheric sensors gather temperature, wind speed, wind direction, wind chill, humidity, solar radiation, pollution factors, rainfall, barometric pressure information, etc. Among the soil characteristics clay content, solar irradiation on the soil, soil's temperature, soil's humidity, pH, salinity, nutrients, organic matter, and soil's depth could be measured.

Optical sensors belong to the category of remote sensing techniques. The optical sensor can be mounted into a portable/static terrestrial sensor, a system of unmanned/manned aerial vehicles and finally satellites. The most common technique is hyper spectral imaging which analyses the complete spectrum of light that reflects from the soil or the plants.

By interface sensors are divided into two main groups:

- **analogue sensors** have advantage of simplicity and low price. Disadvantages - measurements are usually obtained from the sensor in raw form. Measured data have to be recalculated for scaling, zero setting, and linearization calculations are usually done by data recorder (data logger). In case of WSN calculations have to be performed by a microcontroller of node. Disadvantage is that at one signal lead only one sensor could be added. Despite the drawbacks the analogue interface is popular and perspective because each ADC input of modern micro-controller is provided with individually settings for the zero shifts and scaling. Linearization function have to be provided by sensor driver software module;
- **digital or intelligent (smart) sensors** have advantage of scaling, zero setting, linearization, even calibration ability. Digital sensor interface with micro-controller unit is commonly used for connection to addressable data bus, i.e., in a single data line may be a number of sensors.

The most popular sensor buses for the environmental monitoring and precision agriculture are SDI-12, I2C, PWM (pulse wide modulation), SDM (sigma delta modulation), MAXIM 1-wire, UART and some others. The output signals of PWM and SDM sensors are easy convertible by low pass filter to analogue DC voltage output.
By mechanical design sensors can be divided into two groups:

- **constructively finished products** which are connected to the sensor node by cable;
- **sensors built in to the node housing**, usually these are specialized chips. Some of them are used for the monitoring of the node (battery voltage, battery charging and e.t.c., others can be applied for environmental monitoring - solar radiation, ambient light, air temperature and humidity, barometric pressure and e.t.c.

Sensor power consumption. The publications of researches of WSN energy consumption often do not pay due attention to their own sensor energy consumption. Sometimes it can exceed energy consumption of wireless network node. Therefore, it is important to keep in mind sensor energy consumption when selecting necessary sensors, optimize sensor placement, and individual measurement frequency, for example soil moisture measurement may be able to take less frequently than the air temperature measurements.

Results and discussion

WSN Visualization, Monitoring and Managing (VMM) application.
In the framework of this research WSN Visualization, VMM application as a component of WSN prototype with future aim (Fig. 1) for precision agriculture was developed.
Development of VMM application is based on principles described by INSPIRE directive and OGC standards. The standardized approach gives possibility for easy adoption into larger systems in the future.
VMM application deals with:

- monitoring of weather conditions within field equipped with WSN,
- initiation of proactive steps to protect crop from unfavorable weather conditions (e.g. automated cranberry field protection against radiation frost).
- management of historical data according to crop growing environment.
- management of metadata of WSN configuration and sensor observations.

To ensure compatibility in the future with any external system VMM application is built using open interfaces based on Open Geospatial Consortium (OGC) standards and data formats described by INSPIRE directive and the following components:

- Metadata and catalogue system (Micka),
- OGC Web Map Service (WMS), Web Feature Service (WFS) and OGC Sensor Observation Service (SOS) compliant server (MapServer),
- Data storage for vector data (PostgresSQL+PostGIS),
- Data storage for raster data (file system),
- Data maintenance and publishing system (Django, Python frameworks based web application).

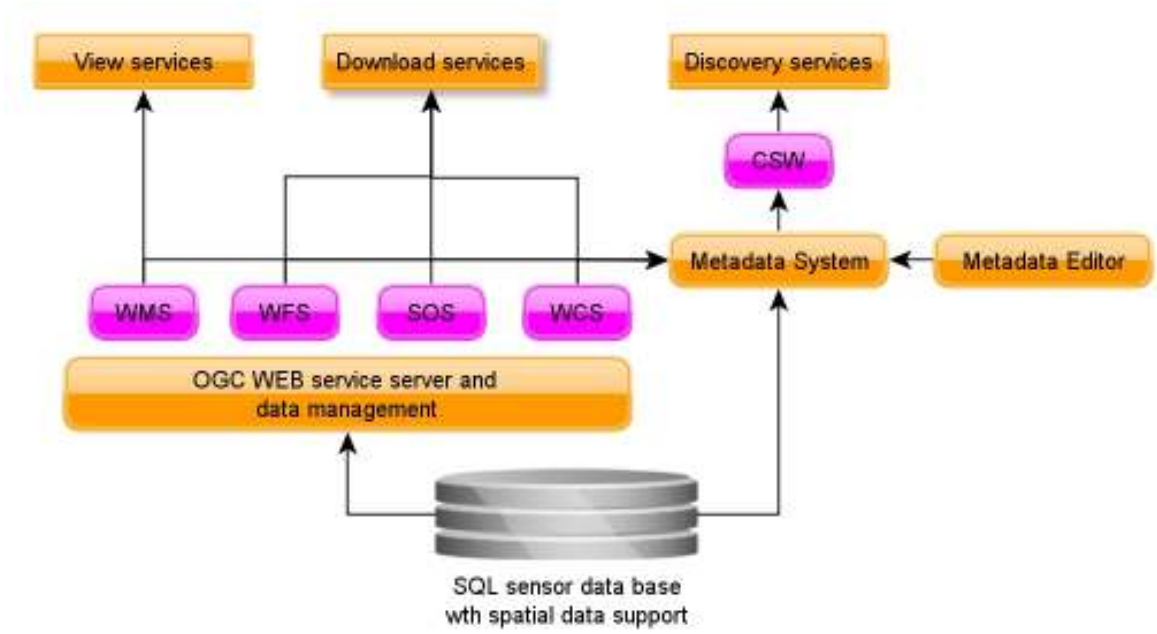


Fig. 2. **Services**.

VMM application is developed using Django, Python, HTML, CSS, AJAX, JavaScript (libraries OpenLayers, GoogleMaps API, GeoExtJS) and it can be accessed by most of modern web browsers.
Postgres SQL is used for sensor data base with open source software plug-in PostGIS supporting spatial geographic objects.
VMM consists of the following modules:
Decision Support Module – (DSM) analyzes observation data from sensor DB by regular polling tables, using mathematical methods (Snyder and Melo-Abreu, 2005) for forecasting dangerous weather conditions and creating event for SMM;
System Managing Module – (SMM) after receiving event from DSM either sends command to Actuator Control Unit (ACU) to start/stop action in **automatic mode** or sends SMS via GSM network to Farmer in **manual mode**.
WSN visualization and monitoring module (WSN VM) allows to view the map of WSN configuration and sensor observation data in different modes depending of Farmer`s settings either in real time or in selected time interval.
VMM allows select **3 levels of access rights**: for administrator (full access), for farmer (view all sensor data and set parameters for SMM and DSM) and for consumer (view all sensor data).
Fig.3 shows curves of both temperature of air at 80cm height and temperature of cranberry leaves during night hours May 13-14 2012. This is excellent example and shows difference of behavior of temperatures. Relative air humidity became as low as 60% without wind and absolutely no clouds. The circumstances reached created conditions typical for radiation frosts.

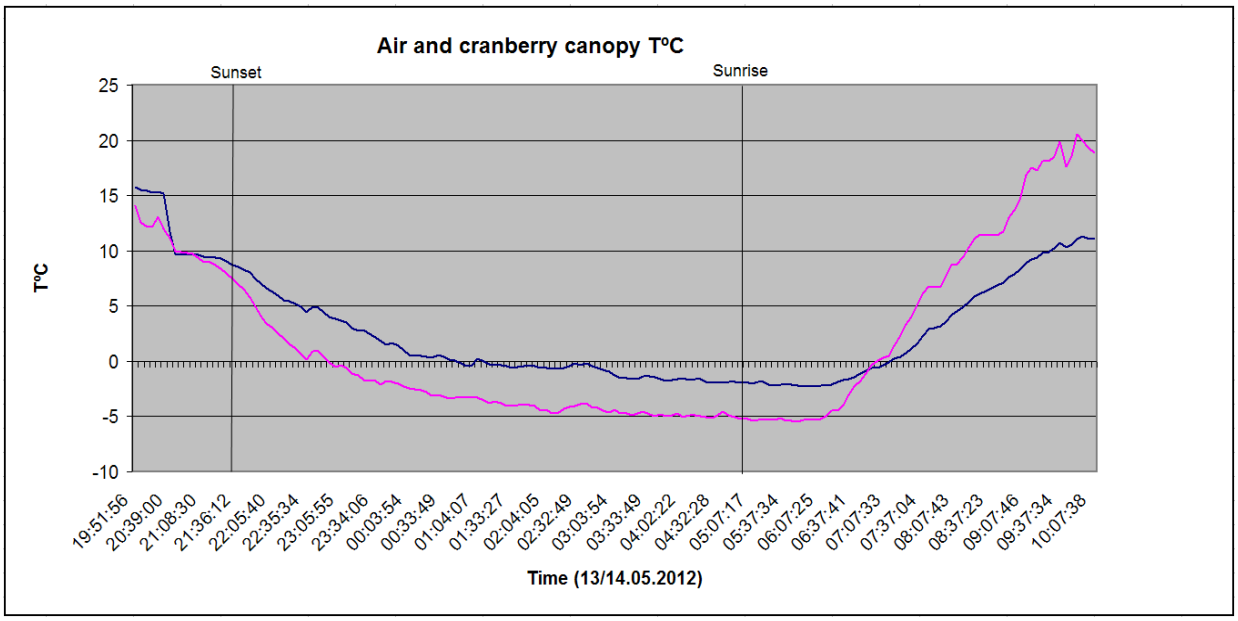


Fig. 3. **Temperature curves**.

During the experiment measurement of temperature, luminosity, air pressure, precipitation, wind strength and direction were collected. The data gathered was necessary input data for using mathematical methods implemented by Decision Support Module (DSM).
Mathematical methods implemented in DSM are based on models of the original paper by Allen for the frost temperature trend model implemented in FFST.xls and FTrend.xls (Snyder and Melo-Abreu, 2005). Two hours after sunset is starting time (t_o) of the model. t_o corresponds to the moment when the net radiation reaches minimum. Assuming there is little or no cloud cover or fog during the night, the net radiation changes little from time t_o until sunrise the next morning. If there are clouds or fog, the model will forecast temperature lower than observed. If a cold front passes or if there is cold air drainage, the forecasted temperature may be too high. For use in the application program for the model, data are selected only from radiation frost nights. Measurements from nights with wind speed values greater than 2.0 ms^{-1} (5 mph) and nights with cloud cover or fog was excluded.
FTrend.xls implemented mathematical model is used to determine temperature trends during a frost night for air, wet bulb and dew point temperatures to forecast how the temperature will change during a radiation frost night from two hours after sunset till sunrise. Sunset and sunrise were determined from the input latitude, longitude and date. The model uses calculations to forecast the air temperature from the start time (t_o) until reaching the forecasted minimum temperature (T_p) at sunrise (at time t_p) the next morning. Air temperature is calculated corresponding to the air and dew-point temperature. Appropriate actuators on the crop field were used when the wet bulb temperature approached to the critical damage temperature.

As mentioned 8], there is another forecast method that was developed by Krasovitski, Kimmel, and Amir (Krasovitski et al., 1996). This method gives more precise forecasting; however implementation of the method is much more complicated. For starting point of researches we decided to apply FAO recommended and above described method. If future researches will show forecasted results will much differ from real ones then it will be useful to try the second method.

As it was discovered during the experiment early forecast is crucial. Because it is too late use any protection, for instance, artificial wind or fog when radiation frost already taking place; the only way is gathering data and calculating probability of radiation frost as early as possible and start protection actions at least 1 – 2 hours before leaves reach dangerous temperature levels.

The experiment gave answer to many questions according to technical realization of WSNs hardware and software. It brought us to idea construct our own based on new concepts WSN hardware and additional alignment of software specifications.

Basic requirements for WSN Operating system (FarmOS)

OS have to be an operating system for large scale WSN. To create that kind of networks there are specific requirements to make network stable. One of the main issues is to guarantee stable communication in long distances with resolution of collisions and self managing topology while meeting legal regulations for radio communications. To support robustness and flexibility of such large scale WSN systems FarmOS have to support:

- collision resolution support for large number of nodes;
- low energy consumption;
- WSN radio communication in three different topologies (star, multihop, mash)
- addressing, MAC protocols, multi-hop routing;
- multitasking;
- unlimited number of software timers;
- compile time configuration for inclusion and exclusion of specific options;
- interactive shell for basic controls and data access;
- integrated development environment;
- analog and digital sensors (including I^2C, SPI, SDI-12 protocols);
- integrated scripting language, etc.

To make system flexible and easy manageable architecture of FarmOS should be structured in three main layers:

- HIL - Hardware Interface layer - hardware independent services that should support any hardware;
- HAL - Hardware abstraction layer - provides services specific to hardware platforms - sets of modules and chips;
- HPL - Hardware presentation layer defines the services by the lowest level of hardware.

On top of the structure there are device drivers and device manager defined allowing unified and selective access to the resources by all applications.

Interface between the user applications and the FarmOS are the system calls.

It would be convenient if FarmOS applications could use plain C and UNIX like system concepts, for instance, sockets for communication. The FarmOS would be easy assessable for most people who have some system programming experience even for those who are new in embedded systems.

As a result an application can be developed without the specification of the hardware. In fact, one of the platforms provided is PC that allows for immediate application debugging and simulation at a high level (Qing Cao et al., 2008).

FarmOS have to have built in scripting language system suitable for common WSN applications in readable and intuitive way. For instance, only a few code lines are required to specify the basic “read sensors” or „send readings to radio” applications.

The graphical user interface allows writing applications in FarmOS even for people with little or no programming experience.

The features above would make FarmOS robust and convenient for agricultural application design, implementation, and maintenance. The main goal is possibility to use large scale WSN systems running on FarmOS in harsh real field environment with little or no ITC personnel support and interaction.

To build large scale WSNs for use in agriculture some specific requirements have been discovered.

Basic requirements for large scale WSN long range node hardware and its solution (Krivanek et al., 2012).

- size of transmitted data (sensor type and the measured value) is approximately 6B (byte) for one variable and one measurement;
- data is read by the Access point node in periodic intervals that can be configured. The shortest possible interval is 60 seconds;
- network nodes must keep working without recharge of batteries at least 6 months (roughly – one agricultural season) performing data collection at least once every 2 hours;
- nodes could be equipped with energy harvesting elements for example with solar cell;

- management and operation of network elements (nodes) must be easy and robust;
- communication range between nodes must be about 250 m to 500 m in the field of forest;
- network nodes should be able to perform simple computational operations; such as zero offset, sensor transfer function, scaling and linearization, dew point calculation from air temperature and humidity measurements e.t.c.;
- easy integration of measured data;
- ability of WSN base station (WSN gateway) to connect to existing public wireless networks;
- operating frequency of 868 MHz or lower in SRD(ISM) bands;
- power consumption of nodes and its sensors must be as low as possible, the capability of recharging its battery for scavenging energy from the environment, and very limited processing capabilities;
- each node and its sensors should conform with enclosures protection ratio code at least to IP65.
- each node should has an unique ID number, written in hardware.

Sensor node should be considered as one entity with its sensors. Without sensors node loses sense of existence. It means that requirements to sensor node mean requirements to sensors connected too. What kind of sensors is needed and what requirements are applied it is defined by application. This research deals with long range sensor network for precision agriculture especially for protection of American cranberry from radiation frosts.

Combining know how obtained from field experiment described above, research in field of WSN, engineering knowledge a state of the art functional schematics of a new (namely - LUMII WSN) node was created (Fig. 4). This schema and mechanical engineering of node solves row of problems of AgroSeNET hardware discovered during research. Schema contains some elements that can be regarded as inventions.

To describe hardware functionality of the LUMII WSN node description should be started with logical components.

Energy harvesting and power management. Different ambient energy sources can produce electrical power. Implementation of photovoltaic energy harvesting today could be done in relatively low costs. Because of its high energy output, photovoltaic cells could be used not only to power up wireless sensor nodes, but also recharge the batteries. Main task of energy harvesting circuit is to gather maximum energy in different solar cell lighting and battery load conditions. For energy harvesting and battery charging of the node solar panel consisting of two solar cells and energy harvesting integrated circuit (IC) LTC3105 from Linear Technologies are chosen. The solution offered is dealing with two tasks; the first one is recharge of the battery, the second is a measurement of solar radiation level. Solar cells are switched in short circuit mode for solar radiation measurement by switching on analog MOSFET switch SW1 (Fig. 4). Solar cell pyranometer – the problem is that it does not measure the whole spectrum of the sun radiation. Therefore it is advisable to use a professional pyranometer to calibrate a photovoltaic instrument.

Two SANYO Eneloop Rechargeable lite long life AAA batteries (up to 2000 cycles) are chosen (Bourgoine, 2011).

Micro-controller. Microcontroller must have enough RAM, to execute the program, and enough flash memory, to store program and data. Microcontroller must have low energy consumption. To connect to all necessary digital sensors and radio transceiver microcontroller must have enough ADC inputs to interface with analog sensors and digital interfaces such as I2C, UART, SDI-12, SPI, 1-wire . The Texas Instruments microcontroller MSP430F2272 with 1 KB RAM and 32 KB flash memory is chosen for LUMII WSN node.

RF module (transceiver). Radio transceiver is the key element of WSN node. It's a main module determining quality of the network. To complete tasks determined by application it is crucial to chose appropriate transceiver chip, frequency band, channel bandwidth, data rate, and modulation technology. As it is mentioned above to perform a measurement of one phenomenon 6 B (bytes) of data are necessary; minimal time interval between measurements is considered 60s for meteorological measurements for agriculture use.

According to European table of frequency allocations and applications frequency range of 863-870 MHz is planned for use of Short Range Devices (SRD). According to ERC recommendation 70-03 relating to the use of SRD devices the maximum transmitting power must be not grater than 25 mW (+14dBm), maximum duty cycle for transmitter is 1%. From these specifications it follows if the time interval between measurements is 60s the transmitter should not be in active mode more than 600ms. Assume, the sensor node measures 8 phenomena – air temperature, humidity, barometric pressure, solar radiation, soil moisture, soil temperature, battery voltage, battery charging current. The common amount of data sent for one measurement cycle to access point, will be 8X6=48B (bytes) or 48X8=384b (bit). To fit in 600ms necessary data transfer rate will be 384bps/0.6s=640bps. Higher data rate requires larger radio channel bandwidth for receiver. According to the Nyquist theorem power of thermal noise applied to the receiver input is proportional to the bandwidth of the channel. The way to increase data transfer rate is by reducing link budget. By increasing data transfer rate data transmitting time and duty cycle decreases hence energy consumption of node goes down. In forested areas multi path fading takes place. Implementing of antenna diversity for forested areas is good solution.

Features of some Texas Instruments RF transceivers, appropriate for low bandwidth, low data rate, long range radio transmission at 868MHz frequency band are shown in Table 1.

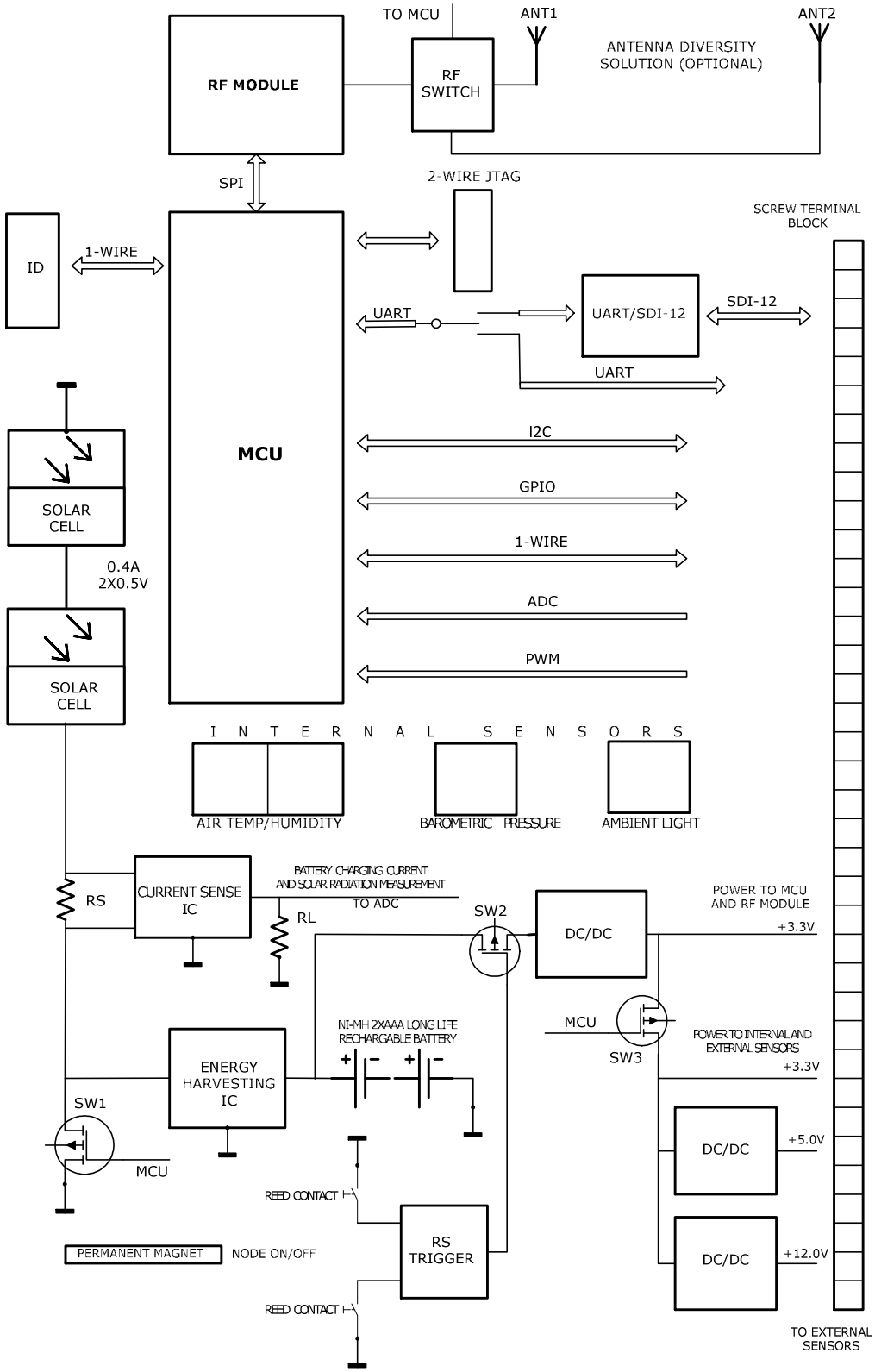


Fig. 4. **Schematics of LUMII WSN node**.

For WSN prototype consisting of 50 nodes Anaren 868MHz RF module based on TI CC1101 transceiver IC has been chosen. Several WSN nodes will be built using 868MHz RF module based on either TI CC1125 or CC1120 transceiver IC will be built for research purposes. Cranberry field protection against radiation frosts requires specific internal and external sensors for new LUMII node.

Table 1

	CC1101	**CC1120**	**CC1125**
Min channel bandwidth, kHz	58	8	2.8
Min channel spacing, kHz	100	12.5	4
Receiver sensitivity at 1200 bps, 1% packet error rate	-112dBm	-123dBm	-123dBm
Receiver sensitivity at 300 bps, 1% packet error rate	N.A.	N.A.	-129dBm
Adjacent channel rejection	37 dB	64 dB	67 dB
Image channel rejection	31 dB	66 dB	66 dB
Current consumption in SNIFF mode	N.A.	3.7 mA	2.0 mA
Average current consumption in RX sniff mode, checking for data packet every 1 second, using Wake-on-radio	N.A.	15mkA	15mkA
Max transmitter output power	+12dBm	+16dBm	+16dBm
Sniff mode	No	Yes	Yes
Wake-on-Radio functionality (WOR)	WOR	Enhanced WOR	Enhanced WOR
Antenna diversity support	No	Yes	Yes
Link budget	124 dB	139 dB	145 dB

Internal sensors:

- air temperature/humidity measurement with dew point calculation possibility - the Sensirion IC SHT21P with PWM interface;
- barometric pressure – the Freescale Semiconductor IC MPXH6115A6 with analogue voltage interface;
- ambient light – the ROHM IC BH1603FVC with analogue current interface;

External sensors:

- soil moisture – the VEGETRONIX VH400 with analogue voltage interface;
- soil temperature – VEGETRONICS THERM200 with analogue voltage interface;
- wind speed - INSPEED VORTEX WIND SENSOR with reed contact output interface (frequency of reed contact switching is measured);
- wind direction - INSPEED E-VANE Hall effect sensor with analogue voltage interface.

Engineering development of the node. The node must be fully protected against dust and heavy rain, i.e., the housing must meet at least IP67 class. The housing of node has to be fitted with transparent cover, to locate the light sensor and solar panel inside the housing. Gas-permeable membrane filter ventilation system must be installed on the housing to provide the same atmospheric pressure, air temperature and humidity of the outside. It gives a chance to fit environmental monitoring sensors inside the housing. This would allow the part of sensors to place within the housing, protecting them from exposure of the environment and reduce the cost of nodes. The junction box is to be used for connecting the external sensors without opening the housing of the node. The cable between junction box and housing of the node has to be pressurized with cable glands. The node is to be equipped with non-contact switch, for example, permanent magnet steered reed, which can be turned on and off. The mechanical design of node has to provide user-friendly and simple fixation on the field.

Conclusion

Deployment of sensor networks still is a problem and subject of wide range researches and developments. Prototype of WSN built in framework of current research shows that small networks are more or less functional while large scale WSNs with long range nodes are issue. Despite of problems our WSN prototypes gave possibility to gather valuable data for field weather monitoring.

The gathered data enabled us to evaluate both conformity with reality of couple of theoretical mathematical models of radiation frost forecasting with reality and features and advantages of AgroSeNET technology in conjunction with just developed WSN operating systems FarmOS beta versions. We got a valuable experience and lessons according to insights of WSNs. Additionally we discovered row of problems using WSNs in great distances more than 300m and in large networks e.g more than 50 nodes.

Acknowledgements

This research has been supported by Supported by European Union via European Regional Development Fund Project “Development of Long Range Wireless sensor network for precision farming applications in Latvia” No.2010/0316/2DP/2.1.1.1.0/10/APIA/VIAA/116

References

Adaptation to climate change in agriculture, forestry and fisheries: Perspective, framework and priorities, 2007. Food and Agriculture Organization of the United Nations, Rome, 24 p.

Agriculture, 2012. Available at: http://www.climate.org/topics/agriculture.html, 15.12.2012.

Bourgoine, N., 2011. Harvest energy from a single photovoltaic cell. *Linear technology journal of analog innovation*. 5 p. Available at: http://cds.linear.com/docs/en/lt-journal/LTJournal-V21N1-01-df-LTC3105-Nathan_Bourgoine.pdf, 15.02.2013.

Goense, D., Thelen, J., Langendoen, K., 2005. Wireless sensor networks for precise Phytophthora decision support, *5th European Conference on Precision Agriculture (5ECP*A), 9 p. Available at: http://www.st.ewi.tudelft.nl/~koen/papers/5ecpa.pdf, 15.02.2013.

Krasovitski, B., Kimmel, E., Amir, I., 1996. Forecasting Earth Surface Temperature for the Optimal Application of Frost Protection Methods. *Journal of Agricultural Engineering Research*. Vol. 63, Issue 2, pp. 93–10. Available at: http://www.sciencedirect.com/science/article/pii/S0021863496900116, 15.02.2013.

Krivanek, Z., Charvat, K., Jezek, J., Musil, M., 2012. *VLIT NODE – new technology for wireless sensor network, ICT for agriculture, rural development and environment-where we are? Where we will go?* CCSS, pp. 245-246. Available at: http://www.ccss.cz/books/ict_agr/ICT_for_Agriculture.pdf, 10.02.2013.

Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, 2012. Special Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*. Cambridge, UK, and New York, NY, USA, 582 p. Available at: http://www.ipcc-wg2.org/, 15.02.2013.

Qing Cao, Abdelzaher, T., Stankovic, J., Tian He, 2008. The LiteOS Operating System: Towards Unix-like Abstractions for Wireless Sensor Networks. Conference: *Information Processing in Sensor Networks – IPSN*. pp. 233-244. Available at: http://www.liteos.net/docs/IPSN_2008.pdf, 01.02.2013.

Snyder, R.L. and Melo-Abreu, J.P., 2005. Mechanisms of Energy Transfer. In: *Frost Protection: fundamentals, practice, and economics.* Vol. 1, Chapter 3. Food and Agriculture Organization of the United Nations, Rome, 17 p. Available at: http://www.fao.org/docrep/008/y7223e/y7223e09.htm#TopOfPage, 02.02.2013.

Snyder, R.L. and Melo-Abreu, J.P., 2005. Frost forecasting and monitoring. In: *Frost Protection: fundamentals, practice, and economics.*Vol. 1, Chapter 5, Food and Agriculture Organization of the United Nations Rome, 16 p. Available at: http://www.fao.org/docrep/008/y7223e/y7223e0b.htm#bm11, 31.01.2013.

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature.* pp. 671-677. Available at: http://www.nature.com/nature/journal/v418/n6898/fig_tab/nature01014_F1.html, 31.01.2013.

Stochastic differential equation approach of height-diameter equations of individual trees

Petras Rupšys[1,2]

[1]*Institute of Forest Management and Wood Sciences, Aleksandras Stulginskis University, Kaunas, LT-53361, Lithuania*
[2]*Centre of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University, LT-53361, Lithuania*
petras.rupsys@asu.lt

Abstract: *In this paper we use a stochastic differential equation to describe the dynamic evolution of the height of an individual tree. The first model is defined by Gompertz shape stochastic differential equation. The second model is defined by Gompertz stochastic differential equation with a threshold parameter. This model can be considered as an extension of the three parameter stochastic Gompertz model with the addition of a fourth parameter. The parameters are estimated by considering discrete sampling of the diameter and height and by using maximum likelihood procedure. Two developed models were employed to compare predicted values with observed values of a height. Performance statistics for developed height-diameter equations included statistical indexes, Shapiro-Wilk test and normal probability plot. We used the data of tropical Atlantic moist forest trees in southeastern Brazil (Scaranello et al., 2012) to validate our modelling technique. Results indicated that our model is able to capture the behaviour of tree height quite accurately. All results were implemented in a symbolic algebra system MAPLE.*

Keywords: diameter, height, mean, stochastic differential equation, threshold parameter, transition density function.

Introduction

In most applications foresters are interested in predicting tree height of a particular tree if only diameter is known. Tree height-diameter relationship establishes quantitative relations between two key characteristic dimensions of trees and is an important component for describing vertical stand structure and for estimating stand volume and site quality. This relationship varies between tree species and stands. Numerous height-diameter mathematical equations of sigmoidal and concave-shaped have been developed using only diameter outside bark at breast height as the predictor variable (Arabatzis and Burkhart, 1992; Rupšys and Petrauskas, 2010c; Petrauskas et al., 2011; Rupšys, 2012; Scaranello et al., 2012). It is well known that many parameters of a tree, including stem density, age, are essential for the dynamic behaviour of the size of a tree height (Hummel, 2000; Ouzennou et al., 2008; Prieditis et al., 2012). Stochasticity may as well play an interesting role in the dynamic behaviour of tree parameters (Rupšys et al., 2007, 2011; Rupšys and Petrauskas, 2010a,b, 2012).

Diameter and height dynamics is affected by many processes and varies among stands. The base assumption of traditionally used regression models is that the observed variations from the regression curve are constant at different values of a diameter would be realistic if the variations were due to measurement errors. Instead, it is unrealistic, as the variations are due to random changes on growth rates induced by random environmental perturbations. Stochastic differential equations models do not have such weakness (Rupšys and Petrauskas, 2012). The modelling of the height-diameter process leads to an equation for the stochastic variable - height, such as a stochastic differential equation, or for an equation which predicts how the probability density function for the height changes in diameter. Stochastic processes are ubiquitous in the physical, biological and social sciences (Allen, 2007).

Stochastic height-diameter dynamics models allow us to reduce the unexplained variability of a height. In recent decades, few models have been put forward to explain stochastic behaviour of diameter and height (Rupšys and Petrauskas, 2010a, b, 2012; Rupšys et al., 2007, 2011). These dynamics are basically the classical deterministic logistic growth dynamics, extended by a level-dependent diffusion term. In reality, external factors such as climate, terrain, the presence of other tree species, and indeed any factor which has an uncertain influence on tree height, will also affect the intrinsic growth rate. This can be modelled by adding an external random term to the intrinsic growth rate, α, which represents this environmental stochasticity. Although many refinements and extensions are possible, the basic dynamics model for height process $H(d),\ d \geq 0$ can be described by the Itô's (1942) univariate stochastic differential equation

$$dH(d) = \mu(H(d),\theta)dd + \sigma(H(d),\theta)dW(d) \tag{1}$$

where $W(d),\ d \geq 0$ is a standard Brownian motion. Intuitively, we interpret the term $dW(\cdot)$ as ecological and environmental noise. Parametric approach assume that the drift $\mu(H(d),\theta)$ and diffusion $\sigma(H(d),\theta)$ are known

functions except for an unknown parameter vector θ. Examples include (Rupšys and Petrauskas, 2010a,b, 2012; Rupšys et al., 2007, 2011). Parametric stochastic differential equations often provide a convenient way to describe the dynamics of tree data, and a great deal of effort has been expended searching for efficient ways to estimate model parameters. Maximum likelihood is typically the estimator of choice (Rupšys et al., 2007; Rupšys and Petrauskas, 2010a, b).

Following the recent trends in stochastic differential equations, we develop a stochastic height-diameter model using the Gompertz shape tree growth. Multivariate models can deal, for instance, with multiple explanatory factors (diameter, stem, basal area) in asset tree height.

This article is an attempt to design a model for predicting the height of a tree from its diameter-varying univariate distribution. For modelling the stochastic process of a height, we use the family of the Gompertz shape stochastic differential equations that are reducible to an Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930).

The aim of this study is to put forward the advantages of using stochastic differential equations in the analysis of height-diameter curves and to show how an adequate model can be made. In this paper attention is restricted to homogeneous stochastic differential equation in the Gompertz type (Rupšys et al., 2007; Rupšys and Petrauskas, 2009, 2010a), whose solution produces the regression term of the fixed effects model. We also discuss how transition density function can be used to construct maximum likelihood estimators. We present an application of stochastic differential equation approach dealing with the study of the height-diameter dynamics of tropical trees. A MAPLE program was implemented to carry out the calculations required for this study.

Materials and methods

We model the dynamic of a height as a stochastic process over diameter. In this study, we select to use the deterministic ordinary differential equation of the Gompertz type (Gompertz, 1825) as the basis of our new developed stochastic height-diameter model. The change in tree height, h(d), is described using the ordinary differential equation

$$\frac{dh(d)}{dd} = \alpha h(d) - \beta h(d)\ln\big(h(d)\big), \tag{2}$$

where α is the intrinsic growth rate of the height and β is the growth deceleration factor. The parameters α and β characterize the evolution of a height of different tree species and stands. The formula describing deterministic Gomperz shape height-diameter trajectory admits the form of a sigmoidal function

$$h(d) = \exp\left(\frac{\alpha}{\beta} - \left(\frac{\alpha}{\beta} - \ln(h_0)\right)\exp(-\beta d)\right),\ d \in [0; D_0], \tag{3}$$

where $h_0 = h(0) = 1.37$. From the solution one can easily see the non-trivial equilibrium point $h(\propto) = \exp\left(\frac{\alpha}{\beta}\right)$ representing the largest height size that a tree can tolerate (i.e. carrying capacity). There also exists an inflection point $h^* = \exp\left(\frac{\alpha}{\beta} - 1\right)$ corresponding to the maximum growth rate of a height, which reflects the self regulation effect by an intrinsic growth control mechanism.

There are alternative ways of introducing stochasticity in the behaviour of a height. In this work, we approximate the randomness in the operation of tree heights as a standard Brownian motion. Therefore, we convert the complete deterministic model defined by Eq.(2) for the tree height into a stochastic model assuming that the intrinsic growth rate varies in diameter according to

$$\alpha(d) = \alpha + \sigma\varepsilon(d), \tag{4}$$

where α is the constant mean value of $\alpha(d)$, σ is the diffusion coefficient, and $\varepsilon(d)$ is a Gaussian white noise process. We describe the height, H(d), using stochastic differential equation of the form

$$dH(d) = [\alpha H(d) - \beta H(d)\ln(H(d))]dd + \sigma H(d)dW(d),\ P(H(0) = 1.37) = 1,\ d \in [0; D_0]. \tag{5}$$

By Ito's lemma Eq.(5) implies that the exponent transformation ψ ≡ ln(h) follows the Ornstein-Uhlenbeck process. This transformation changes the state-space R^+ into R and allows us to obtain the non-conditional probability density function for the considered height process, resulting in

$$f(h,d) = \frac{1}{h\sqrt{2\pi v(d)}}\exp\left(-\frac{1}{2v(d)}\big(\ln h - \mu(d)\big)^2\right), \tag{6}$$

which corresponds with a lognormal distribution, where

$$\mu(d) = \ln(1.37) * e^{-\beta d} + \left(\alpha - \frac{\sigma^2}{2}\right)\left(\frac{1-e^{-\beta d}}{\beta}\right), \tag{7}$$

$$v(d) = \frac{1-e^{-2\beta d}}{2\beta}\sigma^2. \tag{8}$$

The non-conditional mean trend and variance functions of the stochastic height process is given by the following expressions, respectively, by

$$h(d) = \exp\left(\ln(1.37)e^{-\beta d} + \frac{1-e^{-\beta d}}{\beta}\left(\alpha - \frac{\sigma^2}{2}\right) + \frac{\sigma^2}{4\beta}\left(1-e^{-2\beta d}\right)\right), \tag{9}$$

$$va(d) = \exp\left(2\left(\ln(1.37)e^{-\beta d} + \frac{1-e^{-\beta d}}{\beta}\left(\alpha - \frac{\sigma^2}{2}\right)\right) + \frac{\sigma^2}{2\beta}\left(1-e^{-2\beta d}\right)\right)\left(\exp\left(\frac{\sigma^2}{2\beta}\left(1-e^{-2\beta d}\right)\right) - 1\right). \tag{10}$$

Next we propose a new stochastic Gompertz shape height-diameter model with a threshold parameter. This model can be considered as an extension of the three parameter stochastic Gompertz process with the addition of a fourth parameter (Gutierrez et al., 2006; Rupšys et al., 2011). We describe the height, H(d), by stochastic differential equation of the form

$$dH(d) = \left[\alpha(H(d)-\gamma) - \beta(H(d)-\gamma)\ln(H(d)-\gamma)\right]dd + \sigma(H(d)-\gamma)dW(d),\ P(H(0)=1.37)=1,\ d \in [0;D_0]. \tag{11}$$

The non-conditional probability density function for the considered height process (Eq. (11)) is defined in the following form

$$f^t(h,d) = \frac{1}{(h-\gamma)\sqrt{2\pi v^t(d)}}\exp\left(-\frac{1}{2v^t(d)}\left(\ln(h-\gamma) - \mu^t(d)\right)^2\right) \tag{12}$$

which corresponds with a lognormal distribution, where

$$\mu^t(d) = \ln(1.37-\gamma)e^{-\beta d} + \frac{1-e^{-\beta d}}{\beta}\left(\alpha - \frac{\sigma^2}{2}\right), \tag{13}$$

$$v^t(d) = \frac{1-e^{-2\beta d}}{2\beta}\sigma^2. \tag{14}$$

The non-conditional mean trend and variance functions of the height process is given by the following expressions, respectively, by

$$h^t(d) = \gamma + \exp\left(\ln(1.37-\gamma)e^{-\beta d} + \frac{1-e^{-\beta d}}{\beta}\left(\alpha - \frac{\sigma^2}{2}\right) + \left(\frac{\sigma^2}{4\beta}\left(1-e^{-2\beta d}\right)\right)\right), \tag{15}$$

$$va^t(d) = \exp\left(2\left(\ln(1.37-\gamma)e^{-\beta d} + \frac{1-e^{-\beta d}}{\beta}\left(\alpha - \frac{\sigma^2}{2}\right)\right) + \frac{\sigma^2}{2\beta}\left(1-e^{-2\beta d}\right)\right)\left(\exp\left(\frac{\sigma^2}{2\beta}\left(1-e^{-2\beta d}\right)\right) - 1\right). \tag{16}$$

Computation of the parameter estimators

The drift and diffusion parameters α, β and σ are estimated by means of the maximum likelihood procedure using discrete sampling and non-conditional probability density functions (Eqs.(6), (12)), as we assume that all observations (measured trees) are independent. Let us consider a discrete sample of the process (h_1, h_2, . . . , h_n) at the diameters (d_1, d_2, ..., d_n). Under the initial condition $P(H(0)=1.37)=1$, the associate likelihood function can be obtained by the following expressions, respectively

$$L(\alpha,\beta,\sigma) = \prod_{i=1}^{n} f(h_i,d_i) = \prod_{i=1}^{n}\frac{1}{h_i\sqrt{2\pi v(d_i)}}\exp\left(-\frac{1}{2v(d_i)}\left(\ln h_i - \mu(d_i)\right)^2\right), \tag{17}$$

$$L(\alpha,\beta,\sigma,\gamma) = \prod_{i=1}^{n} f^t(h_i,d_i) = \prod_{i=1}^{n}\frac{1}{(h_i-\gamma)\sqrt{2\pi v^t(d_i)}}\exp\left(-\frac{1}{2v^t(d_i)}\left(\ln(h_i-\gamma) - \mu^t(d_i)\right)^2\right). \tag{18}$$

Data

We focus on the modelling of a tropical Atlantic forest tree data set. Here, we analyse tropical forest tree height-diameter database of 280 individual tree height and diameter measurements across plots along an altitudinal gradient published by Scaranello et al. (2012). Our aim is to improve understanding of tropical tree variability and reduce uncertainty of tree height estimates at the altitudinal scale. Summary statistics for diameter outside bark at breast height (D) and total height (H) of all trees used for parameters estimate are presented in Table 1.

Table 1

Summary statistics of the data set (diameter at breast height (D) and total height (H))

Altitude	Count	Variable	Min	Max	Mean	St. Dev.
Sea level	61	D (cm)	4.8	76.9	20.4	15.3
		H (m)	3.0	19.0	10.2	4.2
100 m	73	D (cm)	6.0	75.1	30.5	20.0
		H (m)	4.0	22.0	11.7	4.7
400 m	77	D (cm)	4.9	79.0	30.6	20.7
		H (m)	4.0	25.0	11.3	4.9
1,000 m	79	D (cm)	4.9	100.4	28.6	24.3
		H (m)	3.5	30.0	13.4	6.6
All levels	280	D (cm)	4.8	100.4	27.8	20.9
		H (m)	3.0	30.0	11.7	5.3

Results and discussion

Using the estimation data set presented in Table 1, the parameters of height-diameter models defined by stochastic differential equations (5), (11), were estimated by the maximum likelihood procedure (Eqs. (17), (18)). Estimation results are presented in Table 2.

In general, both stochastic height-diameter models produced relatively high root square mean errors (3.49 m and 3.38 m, respectively) and explained a relatively low proportion of the total variation in observed values of the tree height, accounting for only 57.7% and 60.3%. Nevertheless, these results may not be surprising since the height-diameter relationships found in the data were highly variable and scattered (see Table 1 and Fig. 1).

As we can see in Table 2, the relationship between stem height and diameter are altered by environmental conditions. Stand-specific characteristics such as soil type, nutrient status, elevation cause parameters to differ across stands. Thus, specific stands may have what are generally termed “random parameters” in mixed effects model terminology. Equations (5) and (11) can be altered by adding stand-specific random effects to the population fixed effects parameters to produce stand-specific parameters in the following form

Table 2

Estimated parameters of both models applied to the height-diameter analysis data set

Models	Altitude	Parameters			
		α	β	σ	γ
Eq. (5)	Sea level	0.3872	0.1439	0.1503	-
	100 m	0.3405	0.1236	0.1348	-
	400 m	0.3491	0.1289	0.1456	-
	1,000 m	0.3876	0.1299	0.1386	-
	All levels	0.3698	0.1326	0.1521	-
Eq. (11)	Sea level	0.2908	0.0939	0.0684	-7.0171
	100 m	0.2118	0.0616	0.0368	-14.2301
	400 m	0.2098	0.0628	0.0451	-11.9547
	1,000 m	0.2318	0.0671	0.0496	-10.4416
	All levels	0.2232	0.0644	0.0440	-14.3375

$$\alpha(d) = \alpha + u_{1i}\,, \tag{19}$$

$$\beta(d) = \beta + u_{2i}\,, \tag{20}$$

$$\gamma(d) = \gamma + u_{3i}\,, \tag{21}$$

where u_{1i}, u_{2i}, u_{3i} (i=1, 2, 3, 4) - stand-specific random effects, assumed to be independent and normally distributed with 0 mean and constant variance ($u_{ki} \sim N(0,\sigma^2_k)$, k=1,2,3). Additionally, a covariance, σ_{kl}, can be assumed to exist between u_{ki} and u_{li}.

The influence of the altitude within Atlantic forests on the height-diameter mean and standard deviation curves is illustrated in Fig. 1 and Fig. 2.

For the evaluation of goodness-of-fit of our developed stochastic Gompertz shape height-diameter models (Eqs. (5), (11)) we used the Shapiro-Wilk statistic and normal probability plot. The p-values of the Shapiro-Wilk statistic were 0.005 and 0.195, respectively. The normal probability plots of the pseudo-residuals using the estimates of parameters presented in Table 1 showed that both new developed height-diameter models fit not too bad.

The coefficient of variation is usually used as a measure of precision for the dispersion of data sets and is also often used to compare numerical distributions measured on different scales. The coefficient of variation of a tree height measures the variability of the tree height relative to its mean and it relates the mean and standard deviation by expressing the standard deviation as a percentage of a mean. To further discuss the results of the paper, we provide the coefficient of variation that might to be of considerable interest to compare the dispersion of tree height running at the diameter d, which is defined by

$$CV(d) = \frac{\sqrt{va(d)}}{h(d)} \cdot 100 \tag{22}$$

Fig.3 shows plot of the coefficient of the variation against diameter. In both cases the coefficient of variation of tree height monotonically evolve to stationary coefficient of variation.

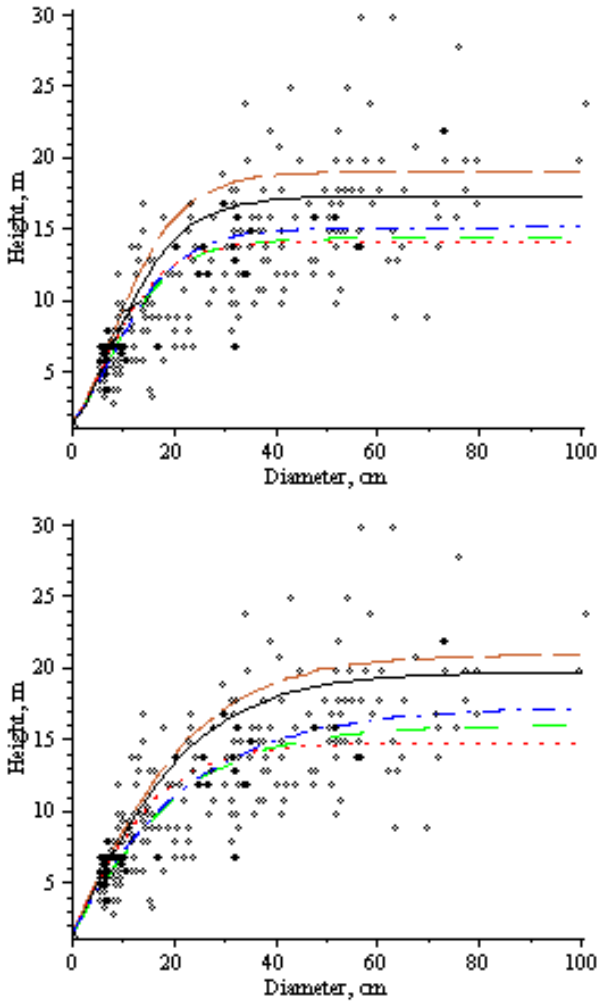


Fig. 1. **Plot of the mean dynamics of a tree height with the parameterization data sets:** in the left – Eq. (5), in the right – Eq. (11), using parameterization data set of sea level altitude – dot (red), using parameterization data set of 100 m altitude – dash dot (blue), using parameterization data set of 400 m altitude – space dash (green), using parameterization data set of 1000 m altitude – long dash (gold), using parameterization data set of all altitudes – solid line (black).

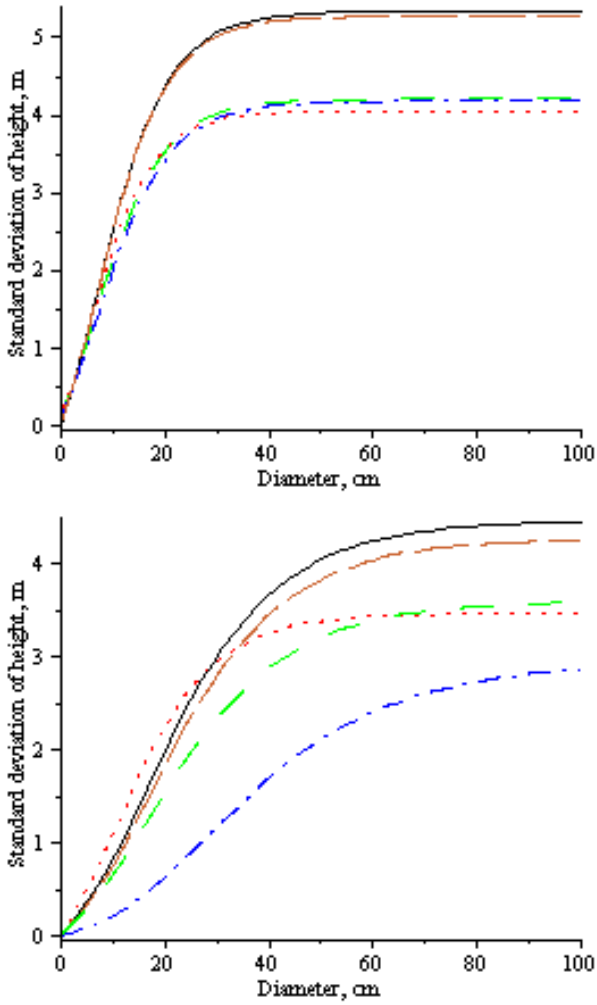


Fig. 2. **Plot of the standard deviation dynamics of a tree height with the parameterization data sets:** in the left – Eq. (5), in the right – Eq. (11), using parameterization data set of sea level altitude – dot (red), using parameterization data set of 100 m altitude – dash dot (blue), using parameterization data set of 400 m altitude – space dash (green), using parameterization data set of 1000 m altitude – long dash (gold), using parameterization data set of all altitudes – solid line (black).

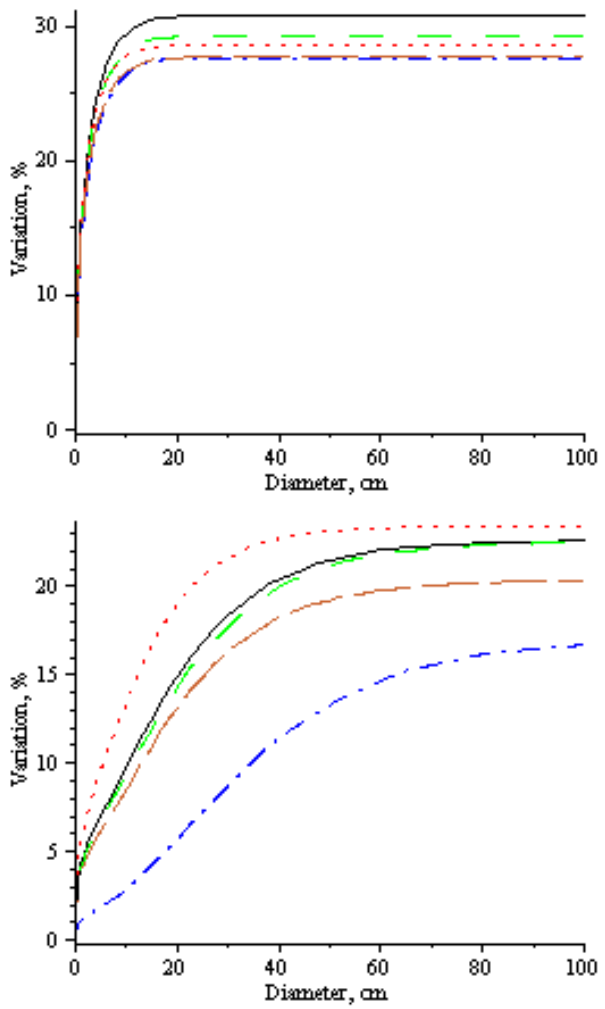


Fig. 3. **Plot of the variation dynamics of a tree height with the parameterization data sets:** in the left – Eq. (5), in the right – Eq. (11), using parameterization data set of sea level altitude – dot (red), using parameterization data set of 100 m altitude – dash dot (blue), using parameterization data set of 400 m altitude – space dash (green), using parameterization data set of 1000 m altitude – long dash (gold), using parameterization data set of all altitudes – solid line (black).

Conclusion

The new height-diameter models were developed using stochastic Gompertz shape differential equations. Comparison of the predicted height values calculated using stochastic differential equations (5), (11) with the observed values revealed a comparable predictive power of the stochastic height model (11).

The developed stochastic models may be recommended both for their ease of fitting procedures and the biological interpretations of the relevant parameters.

The stochastic differential equations approach allows us to incorporate new tree variables, mixed-effect parameters and new forms of stochastic dynamics. The accuracy of the height-diameter-dependent non-conditional density functions (Eqs. (6), (12)) depends on the amount of information available from the stand. Our methodology extends some way to inclusion of the basal-area or/and density of a stand as an exogenous factor or as an independent variable.

The variance functions developed here can be applied generate weights in every linear and nonlinear least squares regression height model by the weighted least squares form.

Finally, stochastic differential equation methodology may be of interest in diverse of areas of research that are far beyond the modelling of a tree height.

References

Allen, E., 2007. *Modeling with Itô Stochastic Differential Equations*, Springer, 230p.

Arabatzis, A.A. and Burkhart, H.E., 1992. An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations. *Forest Science*, 38, pp.192-198.

Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Phil. Trans. Roy. Soc.*, 115, pp.513-585.

Gutierrez, R., Gutierrez-Sanchez, R., Nafidi, A. and Ramos, E., 2006. A new stochastic Gompertz diffusion process with threshold parameter: Computational aspects and applications. *Applied Mathematics and Computation*, 183, pp.738–747.

Hummel, S., 2000. Height, diameter and crown dimensions of *Cordia alliodora* associated with tree density. *Forest Ecology and Management*, 127, pp.31–40.

Itô, K., 1942. On stochastic processes. *Japan Journal of Mathematics*, 18, pp.261-301.

Ouzennou, H., Pothier, D. and Raulier, F., 2008. Adjustment of the age–height relationship for uneven-aged black spruce stands. *Canadian Journal of Forest Research*, 38, pp.2003–2012.

Petrauskas, E., Rupšys, P. and Memgaudas, R., 2011. Q-exponential variable form of a stem taper and volume models for Scots pine (*Pinus Sylvestris*) in Lithuania. *Baltic Forestry*, 17, pp. 118-127.

Prieditis, G., Šmits, I., Archipova I., Daģis, S. and Dubrovskis, D., 2012. Allometric models for predicting tree diameter at breast height. In: Ramos R.A.R., Straupe I., Panagopoulos T. (eds) Recent Researches in Environment, Energy Systems and Sustainability. *8th WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development*, WSEAS LLC, pp.105-110.

Rupšys, P., 2012. On the Use of q-exponential functions for developing stem profile and volume models. *International Journal of Biological Engineering*, 2(5), pp.48-55.

Rupšys, P., Bartkevičius, E. and Petrauskas, E., 2011. A univariate stochastic Gompertz model for tree diameter modeling. *Trends in Applied Sciences Research*, 6, pp.134-153.

Rupšys, P. and Petrauskas, E., 2012. Analysis of height curves by stochastic differential equations. *International Journal of Biomathematics*, 5(5), 1250045. doi No: 10.1142/S1793524511001878.

Rupšys, P. and Petrauskas, E., 2010a. The bivariate Gompertz diffusion model for tree diameter and height distribution. *Forest Science*, 56, pp.271-280.

Rupšys, P. and Petrauskas, E., 2010b. Quantifying tree diameter distributions with one-dimensional diffusion processes. *Journal of Biological Systems*, 18, pp.205-221.

Rupšys, P. and Petrauskas, E., 2010c. Development of q-exponential models for tree height, volume and stem profile. *International Journal of the Physical Sciences*, 5(15), pp.2369-2378.

Rupšys, P. and Petrauskas, E., 2009. Forest harvesting problem in the light of the information measures. *Trends in Applied Sciences Research*, 4, pp.25-36.

Rupšys, P., Petrauskas, E., Mažeika, J. and Deltuvas, R., 2007. The Gompertz type stochastic growth law and a tree diameter distribution. *Baltic Forestry*, 13, pp.197-206.

Scaranello, M.A., Alves, L.F., Vieira, S.A., Camargo, P.B., Joly, C.A. and Martinelli, L.A., 2010. Height-diameter relationships of tropical Atlantic moist forest trees in southeastern Brazil. *Scientia Agricola*, 69, pp.26-37.

Uhlenbeck, G.E. and Ornstein, L.S., 1930. On the theory of Brownian Motion. *Physical Review*, 36, pp.823–841.

ICT IN EDUCATION

Application of IT in Mathematical Proofs and in Checking of Results of Pupils' Research

Juris Cernenoks, Andrejs Cibulis

Faculty of Physics and Mathematics, University of Latvia, Riga, Zellu street 8, Riga, LV – 1002, Latvia
juris2x@inbox.lv; cibulis@lanet.lv

Abstract: *In this paper attention is paid to several aspects of IT applications. Two main aspects: pupils' scientific research in mathematics and computer checking of the results obtained in them. The second aspect is the following one, namely sometimes with the help of a computer in the data generated purposefully it is possible to succeed in seeing a key for nontrivial mathematical proofs. The number of new results having been obtained by a computer researching combinatorial geometry shapes – polyominoes and tetrads have been given. In 2012 several computer programmes for designing of tetrads were developed and, thanks to those results, we managed to find an elegant proof, namely for every n ≥ 11 there is a full polyomino tetrad made of n-omino.*

Keywords: combinatorial geometry, polyomino, pentomino, tetrad, a computer-assisted proof.

Introduction

A polyomino is a connected plane figure formed of joining unit squares edge to edge, *a n-omino* is a polyomino consisting exactly of *n* squares, *a pentomino* is a polyomino consisting exactly of five squares. There are twelve pentominoes named F, I, L, N, P, T, U, V, W, X, Y, and Z respectively, see also Fig.1. The classic reference book on polyominoes is (Golomb, 1994).
Pentomino twins (or *p-twins*) are two equal polyominoes which can be assembled of pentominoes.
A **tetrad** is a plane figure made of four congruent shapes, joined so that each one shares a boundary (of a positive measure) with each. We can find some tetrads as well as information about the first contributors in the Martin Gardner's book (Gardner, 1989). One can find the next contribution in (http://userpages.monmouth.com/~colonel/tetrads/tetrads.html), (http://demonstrations.wolfram.com/Tetrads/). In accordance with (Gardner, 1989): "Michael R. W. Buckley, in the ***Journal of Recreational Mathematics***, 8 (1975), proposed the name **tetrad** for four simply connected planar regions, each pair of which shares a finite portion of a common boundary."
By a *computer-assisted proof* we mean a mathematical proof that has been at least partially generated by a computer. In 1976, the four colour theorem was the first major theorem to be verified using a computer programme.
The combinatorial geometry shapes – tetrads – have some connections with this famous problem, moreover, the very notion *tetrad* appeared almost at the same time when this famous problem was solved. An argument often being made against computer-aided proofs is that they lack mathematical elegance – that they provide no insights or new and useful concepts. In fact, this is an argument that could be advanced against any lengthy proof by exhaustion. Using combinatorial geometry shapes we show that a computer sometimes helps us to find a mathematically elegant proof, thus the mentioned argument is not always valid.
The world of polyominoes is rich in fascinating and easily understandable problems, for which no particular mathematical knowledge is necessary, but which are very far from being easily solved. We were able to obtain complete solutions (or proofs) of some of these problems by means of a computer, and it is unlikely that there could be some other short and easily way. One of the problems of such a type to which we will pay a special attention is the problem of finding the highest pentomino twins (two equal polyominoes which can be assembled of pentominoes). There are two basic guidelines in our selecting of problems: on the one hand, they must be sufficiently difficult or unsolved problems of mathematics, and, on the other hand, they must be understandable and accessible to students or pupils and appropriate as research topics for gifted pupils.

The problem of highest p-twins

The problem – *What are the highest p-twins*? – proposed by Andrejs Cibulis as a contest problem in the Latvian newspaper "Fokus", 1990. Only one solver was able to find twins with the height $H = 16$. Several pupils of Latvia have been investigated pentomino twins problems in their contest papers. Dmitrijs Hromakovs (Form 10, Riga Purvciems Secondary School) was able to find without a computer 227 *little twins* (10-ominoes two copies of which can be constructed from different pentominoes). The computer checking shows that there are 228 *little* p-twins (This pupil did not managed to find only one solution; later he together with his classmate found these last twins). This problem as one of the unsolved problem, appropriate as a topic of pupils' research, has also been proposed in (Cibulis, 2011). After having carefully examined the little twins one can find such twins that are useful as the blocks to construct the highest p-twins, see Fig. 3. A computer analysis shows that the highest

p-twins are only those that can be made of the smaller p-twins, namely with the following structure: (4 + 2), (3 + 3), and (2 + 2 + 2), see Fig. 1-3. Let us mention that there is no uniqueness of building blocks, e. g.:

(FZWY + PL) = (TXUV + NI), (FWUV + TL) = (XYIN + PZ)

(NXW + ZUP) = (YTF + LIV), (FXY + VPL) = (VZN + UTI)

(LW + UY + TP) = (IN + VX + ZF), (LW + UY + TF) = (IN + VX + ZP).

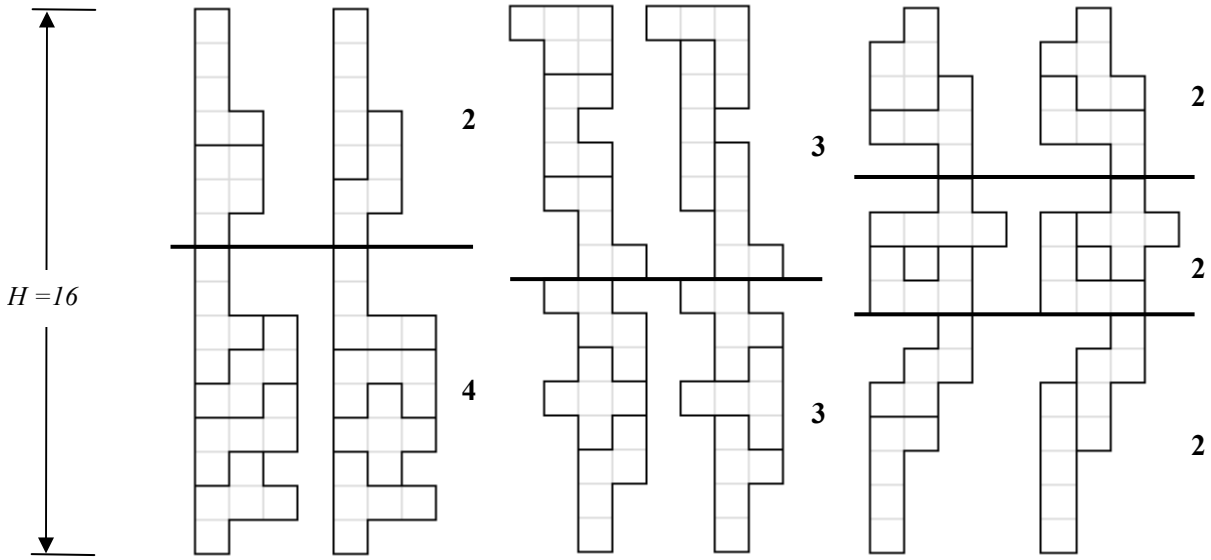


Fig. 1. Fig. 2. Fig. 3.

A short computer-assisted proof that there are no twins with H = 17 is as follows. First of all, 12 pentominoes were divided into two equal size sets. This can be done in 462 distinct ways. Using pentominoes of the one set “towers” were built recursively, in each step adding one pentomino in all possible ways. When "tower" was completed it was immediately tested whether it could be assembled with remaining 6 pentominoes. Few ideas were used to speed up the searching. In each step it is checked whether at least theoretically it is possible to have the height 17. The next optimising related to the order of 6 pentominoes. If the tower was built using pentominoes in some certain order then it is not necessary to check the reverse order, because it gives nothing new. To conclude that twins of the height 17 do not exist the programme worked approximately 6 hours. The idea to use complex numbers in coding polyominoes (see, e. g. (Rangel-Mondragón, 2005)) has been applied to solving this and other problems.

Tetrads

In this section we will formulate and prove the main result of this paper.

The pupils’ first research on tetrads in Latvia was done by Anastasija Jakovļeva in 2012 (Form 10, Riga Secondary School No. 92). She analysed polyomino tetrads. The main aim of her contest paper “Analysis of Tetrads” was to construct tetrads with prescribed properties. For example, find a tetrad having the square n by n as the single hole. To solve this problem it is necessary, and in fact also sufficiently, to find some tetrad with the unit hole. She succeeded in finding such a tetrad made from 10-ominoes, see Fig. 4. The next step is the applying of the well-known idea (each unit square can be replaced by the square 2 × 2), thus we have the so called p r o o f w i t h o u t w o r d s, see Fig. 4-5.

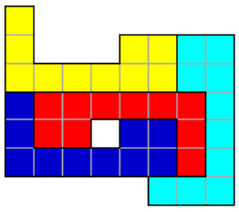


Fig. 4.

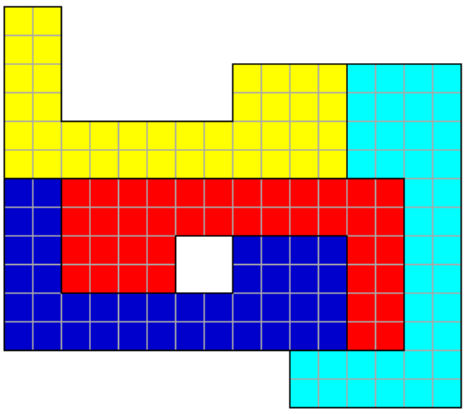


Fig. 5.

The most interesting result in her contest paper is tetrads that are transformable in the square 6 × 6, or in other words, she solved the problem: divide the square 6 × 6 into four equal parts and assemble a tetrad from them. Three partitions of the square are given in her contest paper, see Fig. 6. This result was checked by a computer. It turns out that there are no other solutions.

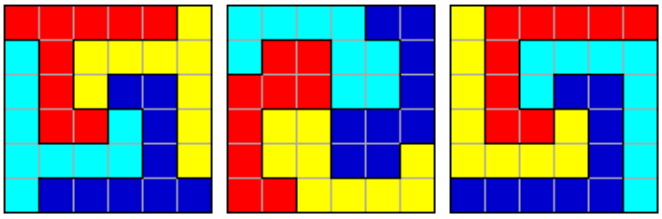


Fig. 6.

The construction of tetrads from these parts of the square would be an easy task and is left to the reader.

R e m a r k . The minimal tetrad having the unit hole has been made from 9-ominoes. There are only two such minimal tetrads, see Fig. 7.

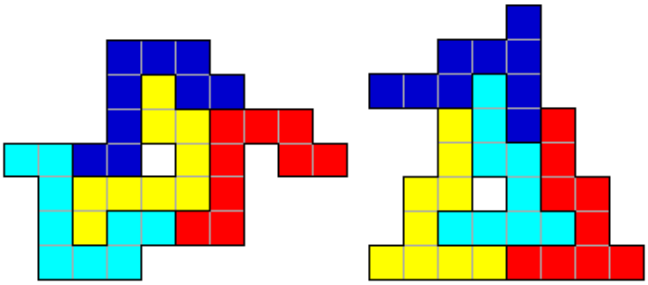


Fig. 7.

A short description of the algorithm for generating tetrads. At the beginning the full list of n-ominoes is generated, and then each polyomino is analysed separately. All the positions of the fixed polyomino are found (there may be up to 8 different positions). One polyomino is fixed. All possible ways to add the next polyomino are detected and they are stored in array. Then all the entries of an array of 3-subsets (the set that contains three elements) are considered. If the 3-subset has a property that every two polyominoes do not overlap and are side by side then the 3-subset together with the early fixed polyomino forms a tetrad. Checking all the 3-subsets can be accomplished with three cycles, but this step is slightly optimised. For example, checking the subsets $\{1; m; n\}$, $\{2; m; n\}$, ..., $\{m - 1; m; n\}$, where $m < n$, it is not necessary to check many times whether m-th and n-th polyomino are not overlapping and adjacent. This step is optimised by using the principle of dynamic programming. In fact, this step is reduced to checking of adjacency and overlapping of each polyomino pair from the array. This algorithm seems to be efficient although several improvements might be made. In addition, the analysis of each pair is the most time extensive, so improvements of this step might bring time savings. Analysing polyominoes that are "far" from each other is redundant. Also more physical memory usage might lead to better results.

The number of tetrads made from *n*-ominoes is given in Table 1. As far as we know the number of tetrads have determined for the first time. Let us mention that computer search for tetrads made from 16-ominoes took approximately 115 hours. Calculations were performed on the oldish HP DV8075 laptop with 2.2 GHz AMD processor and 1 GB RAM. Programme was implemented in Free Pascal.

Table 1

Polyomino tetrads

n	**Number of suitable *n*-ominoes**	**Number of tetrads**
8	8	14
9	42	83
10	187	341
11	739	1388
12	2871	5648
13	11300	22688
14	44440	90243
15	172984	352243
16	670107	1373595

Theorem. For each $n \geq 11$ there is a *n*-omino that forms a tetrad without holes.

The very idea of this theorem is simple. It is necessary to find such a tetrad that can be enlarged to the next one. It is easily to propose such an idea, but it is hard to find such a tetrad. The problem itself of finding small full tetrads is not an easy task, e. g, the *manual* finding of the full tetrad consisting of 11-omino can require several hours of work. Let us emphasise that the key how to prove this theorem was found looking through the huge set of tetrads generated by the computer. The necessary tetrad we found is made from 17-omino.

Proof. The full six tetrads made from *n*-omino, $11 \leq n \leq 16$, are shown in Fig. 8.

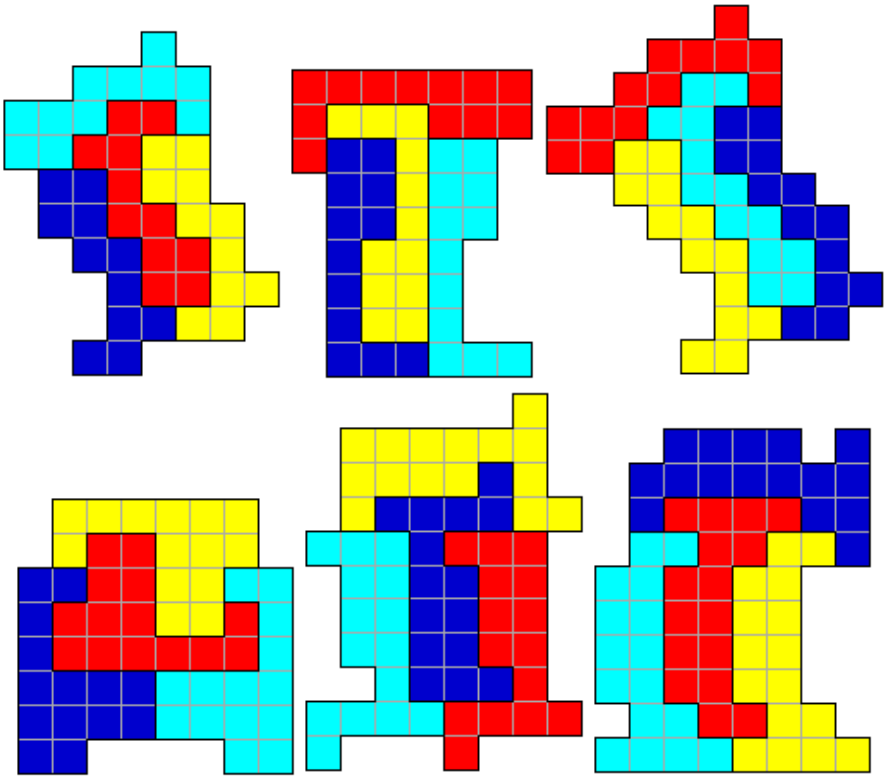


Fig. 8. **Full *n*-omino tetrads for *n* = 11, 12, 13, 14, 15, and 16.**

The key tetrad made from 17-omino is shown in Fig. 9.

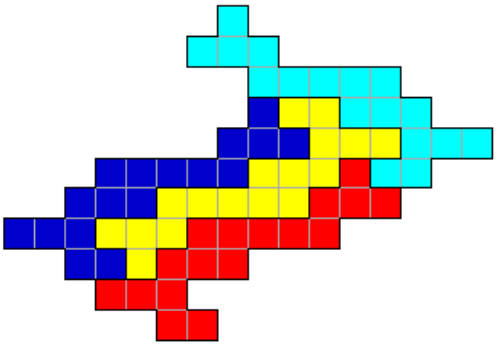


Fig. 9.

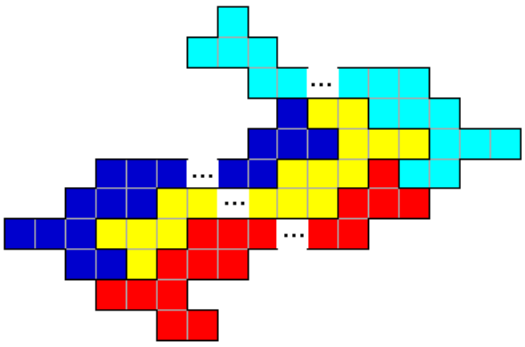


Fig. 10.

Carefully looking at this tetrad one can observe that it can be stretched (extended) arbitrarily long in the horizontal direction, as shown in Fig. 10, thus the theorem has been proved.

Minimal nets

Let $T(n)$ be the smallest number of I-trominoes (1×3) required to stop any more being placed on the board $n \times n$. Each tromino must line up with the squares on the board, so that it covers exactly three squares. The first nine values of the numbers $T(n)$ for $3 \leq n \leq 11$ are as follows: 3, 4, 5, 7, 9, 13, 16, 20, 24. Corresponding nine nets are given in Fig.11.

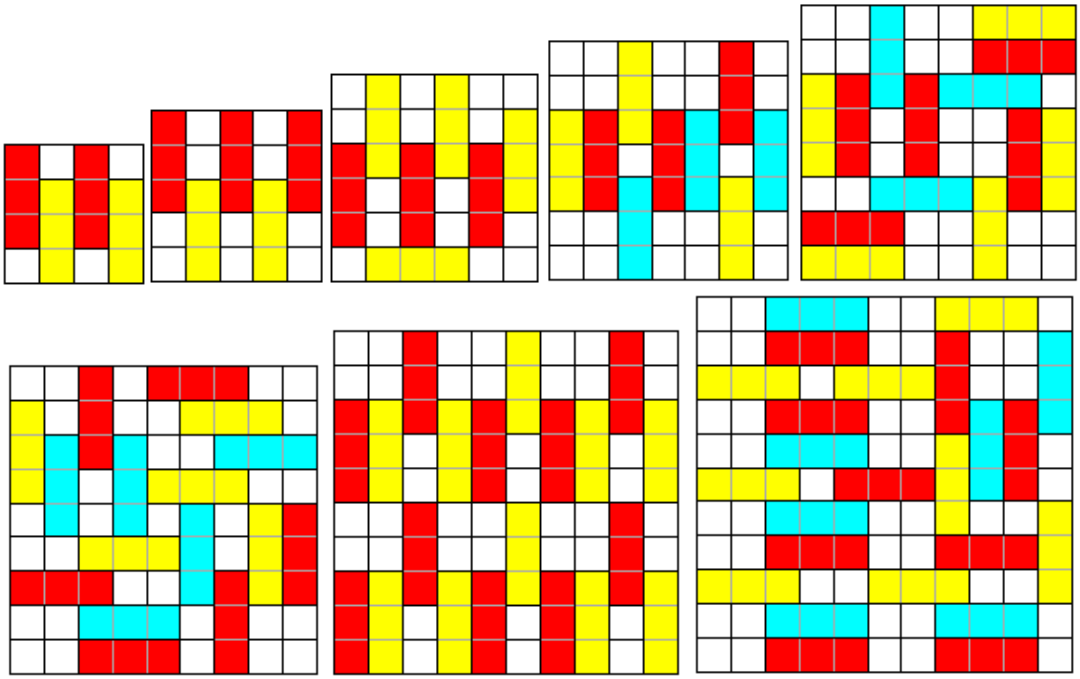


Fig. 11. **Minimal I-tromino nets.**

H y p o t h e s i s : The minimum number $T(n)$ satisfies the inequality

$$\left\lfloor \frac{n^2}{5} \right\rfloor - 1 \leq T(n) \leq \left\lceil \frac{n^2}{5} \right\rceil + 1.$$

Conclusion

The lack of efficient algorithms creates serious difficulties in constructing tetrads. Nowadays a computer plays an important role in mathematics. Firstly, IT applications provide an opportunity to verify pupils' results in scientific research in mathematics. Secondly, a computer allows experimenting and the results give a more profound understanding of the problem and induce new ideas and insights. Since computers become highly sophisticated and more powerful new horizons are revealing constantly. We expect that partially or completely computer-assisted proofs will occur in the future more often. It also has a negative effect, because the programmes are often bulky and difficult to verify. The problem of investigating the numbers $T(n)$ could serve as

the challenging and appropriate topic for pupils' and students' project works and research. This sequence has not been included in the famous Encyclopaedia (http://oeis.org), at least not yet.

References

Cibulis, A., 2011. *Some Bridges among Contest, Research and Unsolved Problems*, The 6th Congress of the World Federation of National Mathematics Competitions, The Proceedings, Riga, Latvia, pp. 54-63.

Gardner, M., 1989. **Penrose Tiles to Trapdoor Ciphers,** Freeman.

Golomb, S. W., 1994. *Polyominoes: Puzzles, Patterns, Problems and Packings*, Princeton, University Press, NJ. (First edition: Polyominoes, New York: Scribners Sons, 1965)

Rangel-Mondragón, J., 2005. *Polyominoes and Related Families*, The Mathematica Journal 9:3, Wolfram Media, Inc.

Evaluation of students drop out reasons in Information Technologies study programs

Liga Paura, Irina Arhipova

Department of Control System, Faculty of Information Technologies, Latvia University of Agriculture, Liela str. 2, Jelgava, LV-3001, Latvia
Liga.paura@llu.lv

Abstract: *The present study analyse the first study course students' dropout rates in higher education institutions, using the real data of Information Technology study program in Latvia University of Agriculture. The problem is to identify, what are the reasons, why only 50% of students completed university and obtained a bachelor's degree. Particular attention was paid to the fact that 30-40 % of students drop out during the first study year. In this research evaluation of the probability of completing University studies was made. Using Survival analysis Proportional hazard model the factors that allow identifying students who are in drop out risk group were described. The following factors were evaluated: students' study duration (month), age, gender, secondary school marks, priority to study in the program (first, second, third) and finance source (budget, private). The results of this study have allowed defining the necessary decision solutions for the 1st study year students' dropout rate decreasing and students motivation increasing to study in information technologies field.*

Keywords: Survival analysis, students' dropout.

Introduction

Survival analysis is a class of statistical methods for studying the occurrence and timing of events and the methodology has been developed over several decades bay Cox, 1972, Kaplan and Meier, 1958. Survival analysis is so named because the method is most often applied to the study of deaths. Although survival analysis was originally developed to analyse cancer data, its use was later extended to study a variety of events (cited by Min et al., 2011)

The analysis of survival experiments is complicated by issues of censoring, where an individual's life length is known to occur only in a certain period of time, and by truncation, where individuals enter the study only if they survive a sufficient length of time or individuals are included in the study only if the event has occurred by a given date. The use of counting process methodology has allowed for substantial advances in the statistical theory to account for censoring and truncation in survival experiments (Klein and Moeschberger, 2005).

Survival analysis is the name for a collection of statistical techniques used to describe and quantify time to event data. There are many different methods used to conduct survival analyses: Life Tables, Kaplan-Meier estimators, Exponential regression, Log-normal regression, Cox proportional-hazards regression.

In 1958, Product-Limit (P-L) method was introduced by Kaplan and Meier (K-M). In the Journal of the American Statistical Association, Kaplan and Meier proposed a way to nonparametrically estimate $S(t)$, even in the presence of censoring (Kaplan and Meier, 1958). The method is based on the ideas of conditional probability and survival function is defined us $S(t) = \Pr(T \geq t)$. $S(t)$ is calculated by Kaplan and Meier estimator:

$$S(t)=\prod(1-(d_t/n_t)), \quad (1)$$

where $t_1, \ldots t_K$ – the set of K distinct death times observed in the sample
d_t –is the number of deaths at time t
n_t – the number of individuals at time t

Kaplan-Meier is technic to analyse survival-time data and to compare two treatment groups on their survival time. Another aim of Survival Analysis is to compare two or more group's survival curves, which usually are made by the Log-rank test (Hosmer and Lemeshow, 1999). The null hypothesis for Kaplan-Meier and Log-rank test is: no difference between the population survival curves (i.e. the probability of an event occurring at the time point t_i is the same for each group).

Cox proportional-hazards regression is useful to identify the risk factors and their risk contributions, selecting efficiently a subset of significant variables, upon which the hazard function depends. Cox's proportional hazards model is analogous to a multiple regression model and enables the difference between survival times of particular groups of respondent to be tested while allowing for other factors (Bewick et al., 2004).

The uses in the survival analysis of today vary quite a bit. Applications now include time until onset of disease, time until stockmarket crash, time until equipment failure, time until earthquake, and so on (Smith and Smith, 2001).

The aim of our study is to evaluate the factors which affect the 1st study course students' dropout rate.

Materials and methods

The data set include 91 full-time students from Faculty of Information technology, enrolled in 2011-2010 academic year at the Latvia University of Agriculture.
Information about students' study duration (month), gender, secondary school marks, priority to study in the program (first, second, third and lower) and finance source (budget, private) were included in the data set.
According the Latvia enrolment rules all potential students may choose several programs at the same time during application process. Students must indicate priority to study in the each program separately (first, second, third etc.) depending on financing source (budget or self-finance). In our case all data have defined by 3 groups: 1st, 2nd, 3rd and lover priority.
Students' dropout was defined, as a student who registered for a study programme, but leaves the University during the first 15 study month. As we can see from the Fig. 1 – 47 students leave the faculty and their study time was between 2 and 12 month.

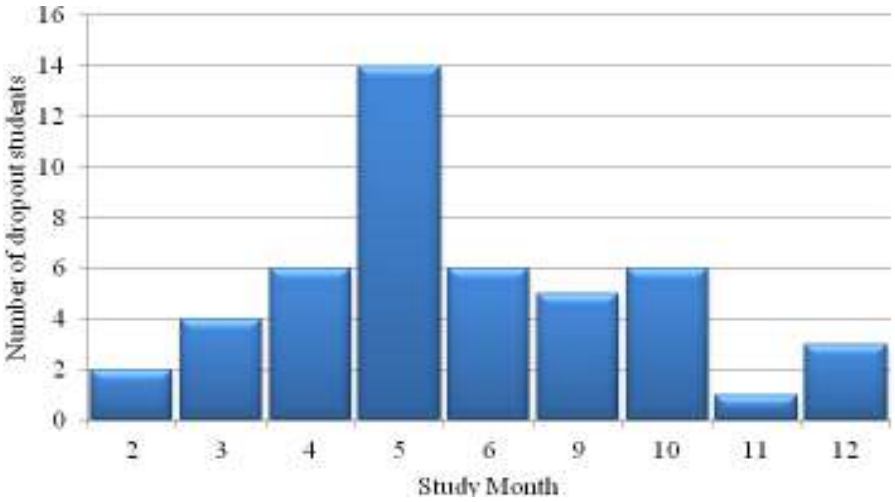


Fig. 1. **Distribution of drop out students study duration (n=47)**.

Study time is a quantitative variable, but the distribution of study time is not normal. Therefore we can't use parametric statistical methods, such as t-test, ANOVA or linear regression for data analysis. Survival analysis methods will be used for study time data and information about censored and uncensored data will be including in the analysis.
Information about censored and uncensored data is reported in Table 1.

Table1

Characteristic of the categorical variables for the whole sample based on number of individuals (n=91)

Variables	n	Study till now, Censored	Dropout, Uncensored
Total	91	44 (48.4%)	47 (51.6%)
Study program			
1 – Cs	61	29 (47.5%)	32 (52.5%)
2 – Pr	30	15 (50.0%)	15 (50.0%)
Gender			
1 – Male	76	34 (44.7%)	42 (55.3%)
2 – Female	15	10 (66.7%)	5 (33.3%)
Priority(n=85)			
1st	56	29 (51.8%)	27 (48.2%)
2nd	8	2 (25.0%)	6 (75.0%)
3rd and lover	21	12 (57.1%)	9 (42.9%)
Mark group			
≤ 25	12	3 (25.0%)	9 (75.0%)
26-35	58	24 (41.4%)	34 (58.6%)
≥ 36	21	17 (81.0%)	4 (19.0%)
Finance			
1 – budget	55	25 (45.5%)	30 (54.5%)
2 – self-finance	36	19 (52.8%)	17 (47.2%)

Study program: 1 – Cs (Computer Control and Computer Science); 2 – Pr (Programming)

Students with a higher probability of dropping out are those who started faculty with school mark 25 and lower (75%) and students with school mark in range 26-35 (58.6%), than students with higher mark. Male students have the highest rates of leaving the faculty.

51.8% and 57.1 % of the students with study priority 1st and 3rd and lover still study at the faculty. Students with the 2nd priority have the highest dropout rate and now only 25% of students from this group are study at the faculty.

For students drop out rate causes the following Survival analysis methods were used:

- Kaplan-Meier was used to compare two groups on their survival time;
- Log-rank test was used to compare two and more groups survival curves and
- Proportional hazard model to determine whether factors influence student dropout.

The proportional hazard model (Cox model) can be written as:

$$h_i(t)=[\, h_0(t)]\, e^{(b0+b1xi1+\, b2xi2+\, b3xi+\, b4x2+\, b5xi5)}, \qquad (2)$$

where $h_i(t)$ – the hazard rate for the ith case at time t
$h_0(t)$ – the baseline hazard at time t
bj – the value of the jth regression coefficient
x_{i1} – gender (1 male, 2 female)
x_{i2} – study programme (1 Cs, 2 Pr)
x_{i3} – finance source (1 budget, 2 private)
x_{i4} – priority to study in the program (first, second, third and lower)
x_{i5} – secondary school marks (covariate)

Factor levels were compared using a levels of significance α=0.05. Statistical analyses were carried out with the program IBM SPSS Statistics 20, IBM, New York, USA (Chan, 2004).

Results and discussion

Kaplan-Meier and Log-rank test

The students drop out rate can be affected by different factors. Each factor independently was analysed by long-rank tests. Summary results of Log-rank test show there are no differences between survival curves for factors study program, gender, priority and finance (Table 2).

Table2

Long-rank test statistics for equality of survival distributions for groups

Variables	Chi-Square	df	Significance
Study program	0.001	1	0.977
Gender	2.013	1	0.156
Priority	4.058	2	0.131
Mark group	11.891	2	0.003
Finance	0.841	1	0.359

Students' secondary school marks were in range 22-48 point. For this analysis the school marks were divided to three groups: 1st mark group ≤ 25 point, 2nd mark group 26-35 point and 3rd mark group ≥ 36 point. There are significant differences between survival curves for factor school mark group ($p<0.05$).

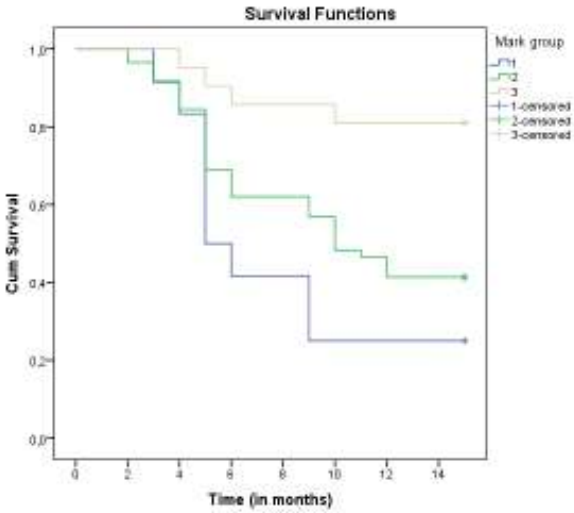


Fig. 2. **SPSS survival plot for factor school mark for three group comparison:** 1st, 2nd and 3rd mark group.

Students who complete their school study with the maximum mark (≥ 36 point), have the higher probability of completing their study, whereas those who have lower marks, between 26 and 35, have the lower probability of surviving, i.e. the lower rate of taking the degree and therefore the higher chance of leaving the faculty (Fig.2).

Proportional hazard model (Cox model)

For presentation the realistic situation when all factors are included to theoretical model the multivariate analysis is more preferable than one factor analysis. For this purpose all factors have investigated together. Main effect model with 5 factors were used for evaluation which factors are significant in students surviving (Table 3). Study program, gender, priority and finance factors were included in the Cox model as categorical covariates (qualitative factor) and school mark as covariate or quantitative factor. As the result finance and priority factors are statistically significant ($p<0.05$) despite to the Log-rank test results. By backward stepwise method the not significant factors study program were exclude from the model.

Table 3

Main effects model by enter and backward stepwise methods (n=91)

		B	SE	Wald	df	Sig.	Exp(B)
Enter method	Gender	0.900	0.561	2.572	1	0.109	2.460
	Study program	0.471	0.386	1.487	1	0.223	1.601
	Finance	1.067	0.374	8.121	1	0.004	2.907
	Priority			5.988	2	0.050	
	Priority (1-2)	1.142	0.477	5.747	1	0.017	3.134
	Priority (2-3)	0.913	0.588	2.409	1	0.121	2.492
	School Mark	-.175	0.047	14.027	1	0.000	0.839
Backward Stepwise method - Step 2	Gender	0.840	0.559	2.254	1	0.133	2.316
	Finance	0.999	0.376	7.084	1	0.008	2.717
	Priority			5.301	2	0.071	
	Priority (1-3)	1.115	0.484	5.293	1	0.021	3.048
	Priority (2-3)	0.650	0.552	1.390	1	0.238	1.916
	School Mark	-.178	0.048	13.864	1	0.000	0.837

Results in Table 3 (step2) show that there are significant differences for finance group ($p<0.05$), priority ($p<0.1$) and school mark ($p<0.001$). A positive sign of coefficient b means that the hazard rate or risk of student's dropout is higher for the first group to compare to the second group and prognosis for that group is worse. Male students, students with budget finance and higher priority are associated with poorer survival, whereas being female; students with self-finance and lower priority are associated with better survival.

The estimated hazard rate for male (coded 1) is exp(0.84) = 2.316 of that of the female; that is, a male dropout risk will be 2,3 times higher than for female after adjustment for the other explanatory variables in the model. However, the p-value = 0.133 is not statistically significant, no difference in survival.
Students with first and second priority are at higher risk than students with third and lower priority to be dropout (HR 3.048, p<0.021; HR 1.916, p<0.238).
A negative sign of coefficient b means that the hazard rate (risk of dropout) is reduced. Students with higher school mark are associated with better survival, whereas students with low mark.
School mark variable the regression coefficient refers to the decrease in hazard rate for an increase of 1 in the value of the school mark. The estimated hazard or risk of dropout decreases by 100%-(100%*0.837) = 16.3% for a one mark unit.
At IT faculty study programme curricula are included such topic us mathematic, physic and chemistry and it is influence the dropout among students. The reasons for students dropout is students' poor knowledge in Mathematics, Physics and Chemistry and poor pre-college academic qualification.
Fig. 3 and 4 displays the estimated survival function of a hypothetical student in interval 0-12 month of study according to different priority (1 to 3) and finance (1, 2). Students are at a very low risk of being dropout at the beginning of the study. The dropout risk slowly increases in the 2nd and 3rd month of study but becomes quite high before the session at the 5th and after session at the 6th month of study.

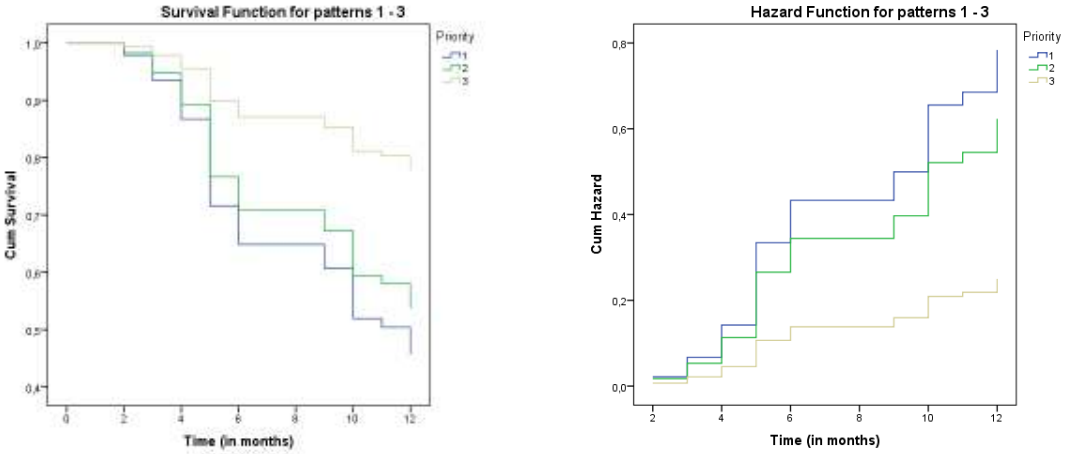


Fig. 3. **SPSS Survival and Hazard plots for factor priority for three group comparison:** 1st, 2nd, 3rd and lover priority.

Therefore, students who start faculty with 1st and second priority have a greater probability of taking the degree than those who decide to study at the faculty with 3rd and lover priority. 27 first priority students (48.2%), 6 second priority students (75.0%), and 9 third and lover priority students (42.9%) are dropping out from the faculty during the first study year. Second priority students have higher hazard curves because, as we have seen from their regression coefficients, they have a greater potential to dropout. During the first 6 month the higher risk to dropout have students with low notes (<25) and second priority.

Table4.

Average school mark in different priority groups
(Standard deviation in brackets)

Priority	Mark - study	Mark - don't study
First	37.5 (6.33)	31.6(3.46)
Second	28.3(0.85)	27.2(3.18)
3rd and lover	28.3(4.99)	27.1(4.47)
In total	34.5(7.22)	30.0(4.18)

The students with first priority have the higher school mark, than students with second and 3rd and lover priority; therefore they have higher risk to drop out. That results show the students with 1st priority and higher school mark leave the faculty (Table 4). The average study duration was 6.9, 4.7 and 6.3 months for the 1st, 2nd, 3rd and lower priority group, respectively.

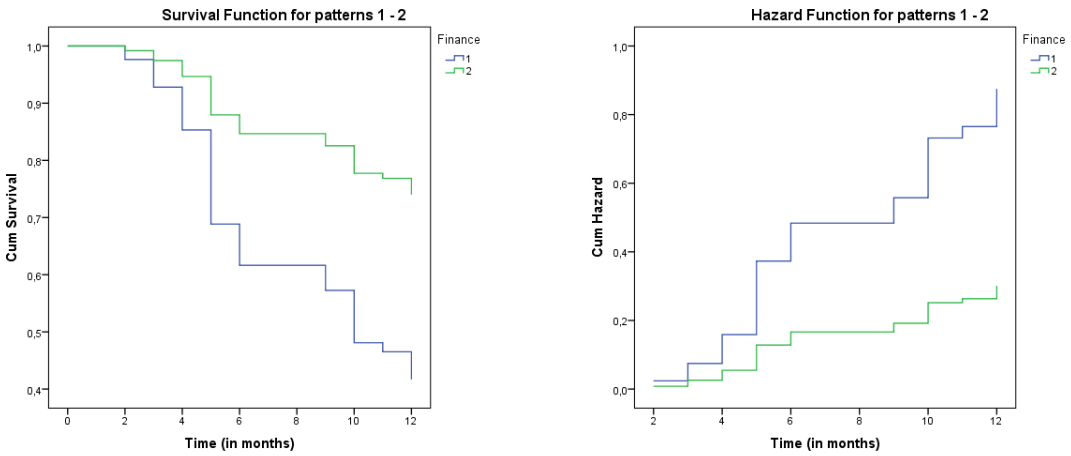


Fig. 4. **SPSS survival and Hazard plots for factor finance for two group comparison.**

According to survival plot can be noted the budget students have a lower survival rate that self-finance students. 30 budget students (54.5%) and 17 self-finance students (47.2%) are dropping out from the faculty during the first study year.

Table5

Average school mark in finance groups
(Standard deviation in brackets)

Finance	Mark - study	Mark - don't study
Budget	36.7 (7.49)	30.8 (4.05)
Self-finance	31.2 (5.78)	27.7 (4.71)
In total	34.3 (7.27)	29.7 (4.50)

The students their finance source gets according to their mark. From Table5 we can see the students, who leave the faculty from the budget group is 30.8 and higher than for self-finance students. The average study duration was for budget students 6.1 months and for self-finance student 6.3 months.
Not only finance factor, priority and school mark can affected the students' dropout rate there are other reason why student leave university.
In literature several factors are associated with student dropout in higher education institutions (Min et al., 2011; Murtaugh et al., 1999): factors associated with attributes or characteristics of the individual student, and factors associated with the institutional environment. When we have students who leave faculty with good mark and from budget, we should analysis the institution environment and topic which studied during the first study year.

Conclusion

1. The results of this study show finance group ($p<0.05$), priority ($p<0.1$) and school mark ($p<0.001$) factors are the main causes for students' dropout at the Faculty of Information Technology.
2. Kaplan-Meier, Log-rank test and Proportional hazard model can be used to evaluate of students' dropout causes.
3. All important factors should be included to the theoretical model for presentation the realistic situation and for this case the multivariate analysis is more preferable than one factor analysis.
4. Data from different study years are recommended to include for further investigations of students dropout rates.

Acknowledgements

Publication of research results has been made due to the funding of the ERAF Project „Promotion of scientific activities of LLU", Contract Nr. 2010/0198/2DP/2.1.1.2.0/10/APIA/VIAA/020

References

Bewick, V., Cheek, L. and Ball, J., 2004. Statistics review 12: Survival analysis. *Critical Care*, 8, pp. 389–394. Available at: http://ccforum.com/content/8/5/389, 18.01.2013.
Chan, Y.H., 2004. Biostatistics 203. Survival analysis. *Singapore Medical Journal*, 45(6), pp. 249-256.
Hosmer, D.W. and Lemeshow, S., 1999. Applied Survival Analysis. *New York – John Wiley and Sons*. 400 p.
Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), pp. 457–481.

Klein, J.P. and Moeschberger, M.L., 2005. Survival Analysis. Techniques for Censored and Truncated Data. *Springer.* 536 p.

Min, Y., Zhang, G., Long, R.A., Anderson, T.J. and Ohland, M.W., 2011. Nonparametric Survival Analysis of the Loss Rate of Undergraduate Engineering Students. *Journal of Engineering Education*, Vol. 100, No. 2, pp. 349–373 Available at: http://www.jee.org. 18.01.2013.

Murtaugh, P.A., Burns, L.D. and Schuster, J., 1999. Predicting the retention of university students. *Research in Higher Education*, 40(3), 355–371.

Smith, T. and Smith, B., 2001. Survival Analysis and the Application Of Cox's Proportional Hazards Modelling Using SAS. *Statistics. Data Analysis and Data Mining.* Paper 244-26. Available at: http://www2.sas.com/proceedings/sugi26/p244-26.pdf , 18.01.2013.

Mathematics e-course for part-time students

Daiva Rimkuvienė, Janina Kaminskienė

Center of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University, Universiteto 10, Kaunas, Lithuania
mfitc@asu.lt

Abstract: *The development of information technologies and the improvement of both students' and lecturers' computer literacy open the possibilities to improve the methodology of teaching Mathematics and to conform to various students' needs. E-learning has already been applied for teaching Mathematics and Statistics course of first- year students at Aleksandras Stulginskis University for several years. This year it has been launched for teaching Mathematics to students of all faculties. The article analyzes the reasons for launching Mathematics course teaching through Moodle environment to AS university part-time students. What is more, it presents both advantages and disadvantages of Mathematics e-teaching. It also presents and analyzes Mathematics teaching material and methodology. Some problems have occurred in the process of creation and administration of Mathematics e-teaching. For instance, these are unfitted tools for evaluation of students' mathematical achievements, lecturers' qualms of efficiency and expedience of Mathematics teaching at a distance, students' insufficient mathematical literacy. One of the advantages is cooperation between lecturers while launching e-courses, i.e. Mathematics topics are analogous for all ASU first-year students only the scope is different, thus lecturers can share teaching material, information, etc.*

Keywords: mathematics, e-learning, Moodle, part-time students.

Introduction

Application of information technologies in teaching has already been analyzed for several decades in the scientific literature. This type of teaching and learning possibilities is now available and attractive for a bigger number of lecturers and students thanks to the improvement of information technology and the development of the Internet connection. Practically, there are no students who do not have skills or technical facilities to use the Internet or modern teaching and communication possibilities. Therefore, virtual teaching environment is successfully applied in various fields of studies. Due to changes in teaching methodology of all subjects Mathematics does not lag behind. There are lots of modern tools that could be used for teaching Mathematics. These are: educational software, video recordings, interactive whiteboards, Intranet, the Internet, etc. The presentation of mathematical symbols does not cause any problem, i.e. text editors and website design tools are suitable for this. Some of lecturers apply calculation software, such as Maple, Mathcad, etc. However, there are discussions ongoing regarding the particularity of Mathematics teaching in e-environment (Marom et al., 2003; Bilbao et al., 2004; Marshall et al., 2012; Vallner, 2012). There are constant doubts whether it is rational, effective and useful or it is just a temporary trend. As the analysis of various literary sources (Marom et al., 2003; Smith et al., 2008,) revealed, teaching Mathematics at a distance is still problematic and not very attractive for both lecturers and students. Firstly, there is a problem of feedback. It is still difficult to check whether a student has solved the task correctly and it is hard to identify where a mistake has occurred because the final result can be written in many different ways (Vallner, 2012). Furthermore, many authors emphasize that distance teaching is successful if a student has motivation for studying. However, the declining literacy can be a barrier that interrupts student's acquisition of knowledge presented in the language of Mathematics. Due to these difficulties Mathematics teaching at a distance is not very common. Nevertheless, changeable habits and skills of using various types of technologies (different to the ones that were popular 20 years ago) forces lecturers to apply information technologies in the process of teaching different subjects, not excepting Mathematics.

Materials and methods

The article is based on the analysis of scientific literature as well as the qualitative analysis of Mathematics e-courses devised at Aleksandras Stulginskis University. Teacher and student surveys have been carried out in order to evaluate the problems that occurred while establishing the e-courses.

Results and discussion

The assumption of Mathematics e-course creation

The problem of the decreasing number of students at our university is due to the complicated demographic situation as well as the ongoing educational reform in Lithuania. Therefore, in order to attract more students the teaching style should be improved and varied facilities should be proposed addressing the diverse needs of students. Our university's students mainly study specialities that are related to agriculture. Thus, possibly, they do not have expectations concerning the studies of Mathematics. Moreover, there are even more problems

regarding the part-time students who not only work but also live quite far from the university. Some of them live not only in the remote regions of Lithuania but also abroad or, for instance, they work in England, Ireland or Norway. Traditionally the studies for part-time students were organized this way: students arrive at the university for a month-length session. During this time they attend intensive lectures, account for their individual tasks and then take an examination at a stated time. During the period between the studies and examinations students are consulted once a week on Thursdays and once a month on Saturdays at the appointed time so that they could arrive and have their questions resolved. What are the problems of organizing this type of teaching and studying? Some of them are:

- Some students have difficulties arranging a month-length holiday from work twice a year;
- It is quite onerous to conceive and soak up the large amounts of information given so intensively;
- It is difficult to arrive at the consultation during the period between the studies and examinations due to the fact that a single-day arrival is inconvenient, expensive and for some of the students even impossible.

Regarding these circumstances the ideas of distance education are being implemented during the recent years. It should be noted that the implementation of the e-courses was not very active; on the contrary, it was carried out thanks to some enthusiastic lecturers. This could have happened because the course implementation encouragement systems were not attractive. The majority of lecturers tended to avoid this type of activity because it was new and required a long time studying the ways and methods of teaching innovatively. However, at the end of the previous school year the university authorities decided that the extended studies should be taught partly in a distance way, paying attention to the implementation of active learning on Moodle platform. Moreover, this should result in a shorter studying session period. Therefore, during this academic year the e-courses of all subjects that are taught to the first-year students of extended studies have been prepared and implemented on Moodle platform.

Students

Several aspects were taken into consideration while preparing Mathematics e-course for particular group of students. These were:

- The level of Mathematical literacy;
- Possibilities and skills of using information technologies;
- Possibilities and skills of studying in a distance education way.

To begin with, it should have been taken into consideration that the majority of part-time students had graduated from schools quite long ago and their Mathematical knowledge was vague. Moreover, some of them had never thought they would ever need this information. Thus the material should have been prepared so that it is easy to understand for a person who is going to renew his/her mathematical skills.

For this reason it is helpful to include not only the new material, but also references where students could find a revision of school topics. Furthermore, for more profound studying references coursebooks should be also included. As far as students' possibilities and skills of using information technologies are concerned, it should be admitted that this does not cause any problems. Currently the Internet connection in Lithuania is good enough. Moreover, there are only rare cases when students do not know how to use a computer. The service of e-mail is widely used for arranging a meeting or giving the tasks or study material. What is more, various study materials and individual tasks are put on the university's website.

However, this could be called a passive teaching, because the student gets information, but there is no active communication or cooperation and the feedback is weak. In order to encourage students to use the virtual learning environment all first-year students are introduced with Moodle system which is quite simple to grasp. If this learning environment is applied in teaching other subjects then it does not take long to acquire a habit of using it. However, there is one thing that bothers – self-discipline and motivation are needed for a successful acquisition of the subject. As the term says itself – it is an e-learning, not the e-teaching system. Thus students should put more effort into studying rather than being passive observers. Though, due to the flexible timetable when students can choose the most suitable time, some of them still have difficulties finding some time at all.

Lecturers

Many years experience has revealed that the best (and the most expensive) way of teaching is an individual teaching, i.e. when a lecturer can give an immediate consultation on student's mistakes and give advice what he/she should learn additionally. A very common and widely used way of teaching is to teach a group of students. It encompasses explanations of task solving methods done on board and during the practice lectures, as well as analyzing how it is going on and explaining the solution to a bigger group of students. It should be taken into consideration that in both cases a student is able to soak only a part of information that is given; especially if the information is given in large amounts and only through the method of explaining. As it is commonly known the acquired knowledge should be revised and students should try to apply it while solving similar tasks. Bearing this in mind the University's Mathematics lecturers are constantly preparing study material so that detailed examples (sources of information) of various ways of solving the tasks were included. Moreover, it is expected

that students would use this material for their additional studies in their free time and speed they wish. It should be noted that students can use not only the material that is prepared by lecturers of ASU, but also the e-books that are prepared by lecturers of other universities. These are available on the University's website. However, the methodological literature that could be found on the Internet is often too complicated to understand for the first-year students because of the barrier of the foreign language. Nevertheless, written sources even if they are presented online are unattractive for first-year students because they do not have the skills of reading and understanding the mathematical information. One of the ways to make learning more diverse is the creation of Virtual learning environment. Various systems could be used for this purpose. These are: Blackboard Learning System, Classroom, Moodle, blogs, Wikis etc. Carr (2000) remarks that students show more favour to those e-courses that are conducted by lecturers who are more skilful in this activity. The experience of teaching Mathematics at a distance was quite poor. Some lecturers (Rimkuvienė et al., 2010) have prepared a small extent course for students of the faculty of Agronomy. Firstly it was designed for working in Blackboard Learning System.

However, the University decided to use Moodle environment and thus the development of this course slowed down as long as lecturers mastered the new environment. Although this did not cause any significant problems, still there were some inconveniences. For instance, each question should have been reviewed and pictures should have been replaced when transferring the questions with pictures from one environment to the other. Moreover, the course presentation structure should have been reorganized, etc.

Advantages and disadvantages

It should be noted that many lecturers had sceptical attitude towards the idea of teaching Mathematics at a distance. Concerning students of engineering specialties the expediency of teaching Mathematics at a distance was most doubtful. During the initial discussions this suggestion was considered to be meaningless and it was even regarded as possibly worsening the quality of teaching. What are the reasons for such approach? Firstly, when teaching students of non-mathematical specialties lecturers face with such problems as poor mathematical literacy and insufficient motivation for studying Mathematics.

We could not agree more with the ideas of Smith G., Ferguson, (2005):

> *Online students often work full time and often have a poorer academic background. Also they are often returning to higher education after a long hiatus. They may have forgotten much of their earlier math skills. This means they often lack the requisite background skills needed for mathematics. Because mathematics is cumulative in nature, with later methods building quite rigorously on earlier methods, mathematics is particularly unforgiving on gaps in background knowledge.*

According to lecturers, when task solution is explained in a face-to-face manner it can be observed from a student's expression when a lecturer has to pause or give a further explanation on something or, for instance, go as far back as school topics. Furthermore, a lecturer can quickly check the tasks that are solved during the practice lectures and identify the character of the main mistakes made. Also, when explaining the solution to one student it can be simultaneously explained to the whole group of students. However, this does not always solve the problem because usually some students are too absorbed into their own tasks that they do not hear the explanation given by the lecturer. As far as recorded material is concerned, it is impossible to get the feedback. However, the recorded material can be watched several times or stopped any time; also it can be played at any convenient time. What is more, distance learning allows placing the answers to questions that are relevant to many students somewhere everybody could check them. When teaching how to solve mathematical tasks it is very important to not only check the final result (sometimes it can be false due to "silly" mistakes), but also the method used and whether the rearrangement was done properly. Unfortunately, e-learning systems are poorly adapted for Mathematics. On the one hand, it is rather difficult to transmit such revision to a computer. On the other hand, computer-based testing is very suitable for checking simpler tasks. Nevertheless any qualms Moodle environment started to be partially used for teaching Mathematics. The model of collaboration was chosen for creating the courses regarding the fact that some of lecturers did not have any experience in using virtual environment for teaching. This is opposed to the model where each lecturer is obliged to create his/her own course. It is obvious that the number of hours for Mathematics lectures and the topics are different for students of different specialities. However, some of the topics are the same. Thus, some parts of the material for one course can be successfully used for the other one. This is relevant when creating the question base. This activity is especially complicated methodologically, intellectually and in terms of time. Having started a large amount of e-teaching it was very beneficial that during the autumn semester Mathematics was taught to part-time students of only one faculty. Therefore, lecturers could pay more attention to identifying the mistakes and improving the course. A large part of the material was used for preparing the course for students of other faculties.

Mathematics course structure and content

According to the schedule the course material was divided into topics. Each topic includes a summary of lectures, explanations, methodological advice on task solution, tests, and individual tasks.

Theoretical material and examples of task solutions are presented in several ways:

- a summary of lectures and methodological advice on task solution – in *pdf* format;

- PowerPoint presentation;
- Video files (slides with recorded explanations; examples of task solutions).

The information that is presented in different formats is not identical. Documents, slides, video files present different examples. Explanations are delivered using various methods. For the purpose of student testing Moodle allows a variety of tools: Multiple choice, Matching, Calculated, Numerical etc. However, not all of the tools are suitable for diverse testing of student's mathematical knowledge. For instance, how could we check whether the calculation of the value of integral is correct? What is more, the tools for writing mathematical formulae on University's Moodle environment (Dragmath Editor, Math&Science by WIRIS) are not user-friendly. Therefore, in order to check students' tasks and their knowledge we use these methods:

- Students take a quiz in Moodle environment;
- Send a photo of a solved task;
- Send a file with a table of calculated answers.

The consultation method depends on the number of students. This year there was a different number of part-time students who chose to study different specialties. The Faculty of Economics and Management admitted 49 part-time students, the Faculty of Forestry and Ecology – 38 students, the Faculty of Land and Water Management – 15, the Faculty of Agricultural Engineering – 10 and the Faculty of Agronomy admitted 25 students. It is obvious that different teaching methods should be applied for a different number of students enrolled in the course. It is possible to work individually with every student if the group is small. However, if the group is big the same method of communication requires more human and time resources. If the number of students is not big then it is possible to communicate more closely and consultations can be organized more often and more careful. If the group is bigger discussions should be encouraged, questions should be generalized and published in the forum.

Consultations were organised in several ways:

If the group was small it was agreed that the lecturer would consult the students once a week via Skype. This way was not acceptable for everybody. There were 5 students who registered for a consultation and it was quite active. The consultations were held both orally and in other ways, i.e. sending messages and files. Other students used to e-mail files that were checked by lecturers and then sent back to students.

Moreover the task hand-in terms and student stimulation methods should be carefully thought over when planning the presentation of a course. This is quite challenging bearing in mind that this course will be used by part-time students who have full-time jobs and have many personal duties.

What is more, time structure in the autumn semester was not very rational. The time period between the sessions was too short to master all the material that was planned according to the program. At the beginning students' performance was quite active. However, later it slowed and even stopped for several weeks and reactivated only just before the beginning of the new session. This was taken into consideration when planning the spring session course. Certain terms were fixed not only for on-line assignments, but also exact dates were appointed for students to arrive at the university and present their midterm tasks.

Conclusion

Although teaching Mathematics through Moodle environment is still complicated and problematic the present demographic and economic situation makes it essential to create Mathematics e-courses with the help of contemporary developing information technologies. Nowadays first-year students do not face big problems using information technologies in study process. However, the lack of mathematical literacy makes it difficult to study Mathematics. Therefore, there is a great need for a constant consequent help during the whole course. Moreover, due to the need of innovative teaching methods and different style of teaching through new teaching environment, lecturers who create e-courses also need help from experienced specialists.

References

Bilbao, J., Bravo, E., Gonzalez, P., Martinez, E., 2004. Teaching Mathematics in University Education through Internet. *Informatics in Education*, Vol. 3, No. 1, p. 19-30.

Boling, E., Hough, M., Krinsky, H., Saleem, H., Stevens, M., 2012. Cutting the distance in distance education: Perspectives on what promotes positive, online learning experiences, *The Internet and Higher Education*,15 (2), p. 118-126.

Carr, S., 2000. As distance education comes of age, the challenge is keeping the students. *The Chronicle of Higher Education*, No. 46(23), A39–A41.

Dias, S.B.; Diniz, J.A., 2012. Blended Learning in Higher Education: Different Needs, Different Profiles. *Procedia Computer Science* vol. 14. p. 438-446.

Marom, R., Chajut, E., Roccas, S., Sagiv, L., 2003. Internet assisted versus traditional distance learning environments: factors affecting student's preferences, 2003. *Computers & Education*. No. 41, p. 65 – 76.

Marshall, N., Buteau, Ch., Jarvis, D., Lavicza, Z., 2012. Do mathematicians integrate computer algebra systems in university teaching? Comparing a literature review to an international survey study. *Computers & Education*. No. 58, p. 423 – 434.

Rimkuvienė, D.; Kaminskienė, J., Venckūnienė, V., 2010. The use of virtual environment for learning mathematics and statistics at the Lithuanian University of Agriculture, *Applied information and communication technologies*: proceedings of the 4-th international scientific conference Jelgava Latvia, April 22-23, 2010 Jelgava. p. 372-375.

Smith, G.G. and Ferguson, D., 2005. Student attrition in mathematics e-learning. *Australasian Journal of Educational Technology*. No, 21(3), p. 323 – 334.

Smith, G.G., Heindel, A.J., Torres, A.T., 2008. E-learning commodity or community: Disciplinary differences between online courses, *Internet and Higher Education*, 1, p. 152–15.

Vallner, H., 2012. Implementation e-learning possibilities in programmed teaching. Teaching Mathematics: Retrospective and Perspectives: proceedings of 13th International Conference Tartu Estonia, 30. May – 1. June, 2012 Tartu. P. 229 – 234.

Predicting students' results in higher education using neural networks

Bogdan Oancea[1], Raluca Dragoescu[2], Stefan Ciucu[3]

[1]*Finance and Accounting Department, "Nicolae Titulescu" University, Calea Vacaresti 185, sector 4, Bucharest, Romania*
bogdanoancea@univnt.ro
[2]*Statistics and Econometrics Department, The Bucharest University of Economic Studies, 6, Piata Romana, 1st district, Bucharest, Romania*
dragoescuraluca@gmail.com
[3]*IT Department, "Nicolae Titulescu" University, Calea Vacaresti 185, sector 4, Bucharest, Romania*
stefanciucu@univnt.ro

Abstract: *A significant problem in higher education is the poor results of students after admission. Many students leave universities from a variety of reasons: poor background knowledge in the field of study, very low grades and the incapacity of passing an examination, lack of financial resources. Predicting students' results is an important problem for the management of the universities who want to avoid the phenomenon of early school leaving. We used a neural network to predict the students' results measured by the grade point average in the first year of study. For this purpose we used a sample of 1000 students from "Nicolae Titulescu" University of Bucharest from the last three graduates' generations, 800 being used for training the network and 200 for testing the network. The neural network was a multilayer perceptron (MLP) with one input layer, two hidden layers and one output layer and it was trained using a version of the resilient backpropagation algorithm. The input data were the students profile at the time of enrolling at the university including information about the student age, the GPA at high school graduation, the gap between high school graduation and higher education enrolling. After training the network we obtained MSE of about 1.7%. The ability to predict students' results is of great help for the university management in order to take early action to avoid the phenomenon of leaving education.*

Keywords: higher education, neural networks, prediction.

Introduction

Romania is among European countries with the highest school dropout rate. The development of strategies to prevent school dropout should be among the priorities for educational institutions. A significant problem in higher education in Romania is the poor results of students after admission. Because of the poor results during the first year of study many students leave universities. The reasons for these poor results are various: poor background knowledge in the field of study, very low grades and the incapacity of passing an examination, lack of financial resources and the increasing costs of education. That's why one of the main aims of the management system of the universities is to avoid the phenomenon of early school leaving. The ratio between graduating students and enrolled students is one of the indicators used for accreditation of higher education institutions in Romania. The universities' management tries to keep this ratio at a good value by means of early intervention and supporting those students with problems and that are candidates to leave school. In this regard, predicting students' results is an important problem and has a great value to universities for early intervention to avoid school leaving.

Many researchers tried to predict the students' results based on various data. Predictions were made using different statistical methods like multivariate regression, path analysis or discriminant analysis. None of these methods have the power of discovering potential data patterns as neural networks. Feed forward neural networks are applied in many fields like financial forecasting, medical diagnosis, bankruptcy prediction, OCR for regression or classification purposes because they are one of the best functional mappers. The good results of applying neural networks in classification problems lead us to use them for predicting students' results in higher education.

In this paper we describe an implementation of a feed forward neural network used to predict the GPA after the first year of study. The paper is organized as follows. The next section reviews the related work in the field of using neural network for prediction and classification purposes in education. Then we describe our implementation of the neural network and the data set used for training and testing. Next, the results of using the network are presented, showing the good capabilities for classification of the NN. The final section presents some conclusions and ideas for further development of the study.

Related work

Neural networks were used by many researchers for predictions of students' results. In (Cooper, 2010) the author presents a neural network-based decision support system that identifies students who are "at-risk" of not retaining to their second year of study. The system correctly predicted retention for approximately 70% of the students. (Halachev, 2012) presents a neural network used for prediction of the outcome indicators of e-Learning, based on Balanced ScoreCard. The author obtained a 3-4% prognosis error which is acceptable from a practical point of view.

(Livieris et al., 2012) describes the implementation of a user-friendly software tool based on neural network classifiers for predicting the student's performance in the course of "Mathematics" of the first year of Lyceum. The authors of the paper compared many training algorithms namely the Broyden-Fletcher-Goldfarb-Shanno (BFGS), the Levenberg-Marquardt (LM), the Resilient Backpropagation (Rprop) and the modified spectral Perry (MSP). The performances of the neural network classifier were also compared with other classifiers such as Bayesian networks and support vector machines and the authors proved the better results of the neural network classifier.

Data registered by Moodle were used in (Calvo-Flores et al., 2006) for predicting students' marks. The authors used a RBF neural network and information logged by Moodle regarding the number and types of educational resources accesses to predict the students' marks at a discipline. They obtained an accuracy of prediction of about 75-80%. The model developed in this paper shows that it is possible to predict those students with problems to pass a course and to give professors an indication to pay more attention on those students that probably will fail passing a course.

In (El Moucary et al., 2011) the authors present a Neural Network and Data Clustering based method designed to predict students' GPA according to their foreign language performance for those students who study in a foreign language. In a second stage the students are grouped in well-defined clusters for further advising. They obtained a maximum error of prediction less than 10% for GPA.

Neural networks were also used in (Naik et al., 2004) to predict MBA students' success. The authors classified applicants to MBA program into successful and marginal students pools based on undergraduate GPA, undergraduate major, age and GMAT score using a neural network with three layers. They obtained a prediction accuracy for their model of about 89%. To assess the ability of the neural network for classification of students, the authors compared the results obtained using the neural network with a logit and probit regression model. The overall rate of prediction accuracy of the logit model was about 73% while the accuracy of the probit was 73.37%.

(Oladokun et al., 2008) used a neural network in order to determine those factors that influence students' performance. They classified students in three classes according to their results. The accuracy of prediction obtained by the authors of the paper was about 74%. (Karamouzis et al., 2008) used a three-layer perceptron network trained by backpropagation to predict student graduation rate. The network model build by the authors had an accuracy of 70.27% for successful graduates and 66.29% for unsuccessful graduates.

Methodology

For our study, we used a sample of 1000 students from the last three graduates' generations from "Nicolae Titulescu" University from Bucharest. Most of the students that leave studies take their decision after the first year. In our data sample, about 70% of the students who left university took this decision after the first year of study. Most of them had a very low GPA, usually lower than 6 (in Romania, grades range from 1 to 10). In order to avoid the phenomenon of early school leaving, we build a model to predict the GPA of the students after the first year of study and classify them in three classes according to theirs GPA. A lower GPA is a serious indication that the student has difficulties in finding his/her educational path and is a premise for early school leaving.

Our model uses a neural network with one input layer, two hidden layer and one output layer. As input data for predicting the GPA after first year we used:

- Type of the study program: distance education (part time) or full time education;
- Gender of the student;
- High-school graduation GPA;
- Age of the student;
- Difference in years from the moment the student graduates high-school until he/she enrolls at university.

We classified students according to theirs GPA after the first year of study in three classes:

- POOR RESULTS – those students with GPA lower than 6;
- MEDIUM RESULTS– those students with GPA between 6 and 8;
- GOOD RESULTS – those students with GPA greater than 8.

Our task was to predict the class a student belongs based on the five input variables. The first layer of the neural network comprises 7 neurons: two for type of education (0/1 full time/part time), two for gender (0/1 – M/F), one for high school graduation GPA, one for the age of the student and one for the difference in years from the moment the student graduates high-school until he/she enrols at university. The last three variables were normalized to [0, 1] interval. We conducted a series of tests in order to establish the number of hidden layers and the number of neurons in each hidden layer. Our tests give us that the best results are obtained with two hidden layers, the first layer having 50 neurons and the second layer having 400 neurons. The output layer has three neurons, one for each of the three classes. The input layer and each of the hidden layers also receive a bias signal of 1. Every neuron in a layer is fully connected to every neuron in the next layer.

Each neuron accumulates the input from the neurons in the preceding layer and calculates an output signal according to:

$$y_i = f(\sum w_{ij} x_i + b) \tag{1}$$

where b is the bias input and f is the activation function of the neuron. We used the *tanh* as an activation function for the two hidden layers and *softmax* function for the output layer. We used MSE (mean square error) as a network error function.

We tested several training algorithms: backpropagation, quick propagation, classical resilient propagation, scaled conjugate gradient. We find that the best results are obtained using a version of the Resilient backpropagation called iRPROP+ which is one of the best performing first-order learning methods for feed forward neural networks (Igel, 2000).

The classical RPROP training algorithm controls the weight update for each connection thus maximizing the update step size and minimizing oscillations. The direction of weight update is based on the sign of the partial derivative $\frac{\partial E}{\partial w_{ij}}$, where E is the error function and w_{ij} is the weight from neuron j to neuron i. The update size is different for each weight and is independent of the absolute value of the partial derivative. If the partial derivative $\frac{\partial E}{\partial w_{ij}}$ has the same sign for consecutive steps, the step-size is increased, otherwise it is decreased.

The weight updates are computed according to the change in sign of the partial derivative: if the sign has not changed, the weight update is done normally but if the sign has changed the previous weight update is reverted (Riedmiller, 1993).

The iRPROP+ training algorithm modifies the rule of weight updating and reverts only weight updates that have caused sign changes of the partial derivate and an error increase. This rule combines information about the sign of the error function derivative which is error surface information with the magnitude of the network error when the decision of reverting an update step is taken.

Results

We have implemented the neural network with the Encog framework (Heaton, 2011) using Java programming language. We used a sample of 1000 records from the last three graduates' generations: 800 records where used for training the network and 200 records were used for testing the network.

Choosing number of the neurons and layers is a difficult problem. A small number of neurons and layers will lower the mapping power of the network. On the other hand, a large number of hidden layers and neurons could give the network the power to fit very complex data but it will slow down the training process. The network structure was found on a trial and error basis. We started with a small network and gradually increase its size. Finally we found that the best results are obtained for a network with the following structure: 7I-50H-400H-3O, i.e. 7 input neurons, first hidden layer with 50 neurons, a second hidden layer with 400 neurons and an output layer with 3 neurons.

We trained the network for 100.000 epochs on a computer with INTEL I7 processor and 4 GB of RAM memory under the Windows 7 operating system. Before training the network, the order of the training data records was randomized. The MSE obtained after training the network was 1.7%. The mean square error for the test data set was 1.91%.

In our data set 30.1% of students belong to the class "POOR RESULTS", 50.9% of them to the class "MEDIUM RESULTS" and 19% to the class "GOOD RESULTS".

Table 1 shows the number of students in each class predicted by our neural network using the test data set compared with the real number of students in each class.

Table 1

The results of the neural network predictions

CLASS	Number of students	Predicted Values	Predicted value (percent)
POOR RESULTS	60	52	86,6%
MEDIUM RESULTS	105	99	94.2%
GOOD RESULTS	35	30	85,7%

The results presented in table 1 are encouraging: 86.6% of the poor results students were predicted by the network.

Conclusion

Prediction is an important tool and represents a first step in intervention from the university management to avoid the phenomenon of early school leaving. We used the classification power of a neural network to predict the potential students with problems in continuing their education. Our network achieved an accuracy of over 86%.

We used the Encog framework for building the network that is a feed forward Multi Layer Perceptron with one input layer, two hidden layers and one output layer. The activation function of the hidden layers is *tanh* and for the output layer we used the *softmax* function. We tested several training algorithms and we found out that the best results were obtained using the iRPROP+ algorithm.

The average predictability rate was 86% for the "POOR RESULTS" class of students which represents the pool of potential students that are candidates for leaving the university. This predictability rate is comparable with other results presented in the beginning of the paper. This encourages us to continue the research, first of all by expanding the number of input variables included in each student' record. A larger number of input variables will increase the predictability power of the network. Another improvement would be to increase the number of training records which is also useful for cross-validation.

Another direction for a future research would be to identify those parameters from the input data that influence in a greater extent the students' decision to leave education. This will be helpful for educational administrators to intervene in the shortest possible time. Eliminating these parameters will contribute to an increasing predictability power of the network since they are only noise.

References

Baker, B.D. and Richards, C.E., 1999. A comparison of conventional linear regression methods and neural networks for forecasting educational spending, *Economics of Education Review,* 18, pp. 405-415.

Calvo-Flores, M.D, Galindo, E.G., Pegalajar Jiménez, M.C. and Piñeiro, O.P., 2006. Predicting students' marks from Moodle logs using neural network models, A. Méndez-Vilas, A. Solano Martín, J.A. Mesa González and J. Mesa González (*eds*.) Current Developments in Technology-Assisted Education, pp. 586-590.

Cooper, C.I., 2010. A Generalizable Neural Network for Predicting Student Retention, Professional Papers Proceedings Archive, Available at: http://www.asee.org/documents/zones/zone1/2010/professional/A-Generalizable-Neural-Network-for-Predicting-Student-Retention.pdf, 17.01.2013.

Halachev, P., 2012. Prediction of e-Learning Efficiency by Neural Networks, *Cybernetics and information technologies*, Volume 12, No 2, pp. 98-108.

Heaton, J., 2011. Programming Neural Networks with Encog3 in Java, 2nd Edition, Heaton Research.

Igel, C., Hüsken, M., 2000. Improving the Rprop Learning Algorithm, *Proceedings of the Second International Symposium on Neural Computation*, pp. 115–121, ICSC Academic Press.

Karamouzis S.T., and Vrettos, A., 2008. An Artificial Neural Network for Predicting Student Graduation Outcomes, *Proceedings of the World Congress on Engineering and Computer Science*, pp. 991-994.

Livieris, I.E., Drakopoulou, K., Pintelas, P., 2012. Predicting students' performance using artificial neural networks, *Proceedings of the 8th Pan-Hellenic Conference "Information and Communication Technology in Education*", Available at : http://hcicte2012.uth.gr/main/sites/default/files/proc/Proceedings/LivierisEtAl.pdf, 17.01.2013.

El Moucary, C., Khair, M., Zakhem, W., 2011. Improving Student's Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education, *The Research Bulletin of Jordan ACM*, vol II, pp. 27-34.

Naik, B., Ragothaman, S., 2004. Using Neural Network to Predict MBA Student Success, *College Student Journal*, Vol. 38, No. 1. pp. 1-4.

Oladokun, V.O., Adebanjo, A.T., and Charles-Owaba, O.E., 2008. Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, *The Pacific Journal of Science and Technology*, Volume 9, Number 1, pp. 72-79.

Riedmiller, M. and Braun, H., 1993. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 586–591. IEEE Press.

Preparation of the mathematics teachers of the Republic of Kazakhstan in the conditions of informatization of education

Bakhytkul Kaskatayeva

Kazakh National Peadagogical University Abai, 13 Dostyk ave., Almaty, 050010, Republic of Kazakhstan

Abstract: *The paper considers the preparation of the future teachers of mathematics to the professional activities in the conditions of informatization of education. The model of training the future teachers of mathematics to a professional activity is analyzed. The model is characterized by the integrity of the structures formed a unity of purpose, organizational principles, content, forms and methods of training. The potential for learning using ICT is analyzed the creation of real contexts for (1) learning the multiplicity of types of visual presentation, (2) encourage the active and reflexive learning; (3) increase of the graphic skills and (4) increasing the effectiveness of time consumption for the solution of the problem.*

Keywords: training, education, mathematics, methods of teaching, informatization.

Introduction

One of the acute problems of the modern higher school is the problem of preparation of the future teachers of mathematics to the professional activity in the conditions of informatization of higher education. In accordance with the Concept of education development of the Republic of Kazakhstan for the period up to the year 2015, and with the Strategy of informatization of the educational system of the Republic of Kazakhstan till the year 2020, in our Republic are determined by the basic directions of activity on introduction of modern information technologies.

The modern information society puts forward new requirements to the training of education in matters of use and the production of an information resource of the Internet, development of methods and means of informational interaction in the local and global networks, realization of opportunities of information and communication technologies (ICT) in the process of self-recovery and knowledge representation. The demand for teaching staff, able to work in conditions of the use of distributed information resource information networks, able to take responsibility for the realization of the opportunities of ICT in their professional activities, ready to continually improve their professional level adequately modern trends in the development of the information society.

However, in the practice of teaching of higher mathematics in the modern education school these issues are not always given due attention. The computer becomes more accessible, the Internet has a wide distribution in the entire world. Through the Internet the vast amount of information becomes accessible, but we need to teach students to receive this information, assess its quality and use correctly. Information communication technology forms a new perspective on learning and teaching. Learning can no longer be viewed as a process in the framework of the lecture notes and textbooks. Many studies show that ICT is an excellent tool that contributes to the formation of new ideas about the processes of teaching and learning.

The decision of problems of preparation of the future teachers of mathematics to the professional activity, as it seems to us, is in the creation of a fundamentally new didactic system or learning model, implying the optimal informational interaction of the teacher and graduate students, as well as a master student and computers in educational process of higher education institution (University).

A model of training the future teachers of mathematics to a professional activity under the conditions of informatization of education is presented and analyzed in this study.

The model of training the future teachers of mathematics

Under the didactic system of means allotted by the certain criteria of holistic education didactic systems and models are characterized by integrity of structures formed a unity of purpose, organizational principles, content, forms and methods of training. We have developed a model of training the future teachers of mathematics to a professional activity (Fig. 1) and an electronic textbook on the theory and methods of teaching mathematics (Bent and van den Brink, 2005; Kaskatayeva, 2009).

The main aim of our research is the development of methods of preparation of the future teachers of mathematics to a professional activity under the conditions of informatization of education and the definition of the forms, methods and means for its implementation.

It's very important for the solution of this problem is the attitude of teachers to the informatization of the educational process.

The results of the pilot study in our Institute of magistracy and PhD studies at KazNPU named after Abai. The application of the information technology maintenance of educational process on a number of cycles of

disciplines show, that already today in the conditions of informatization of education significantly change the role, place and tasks as a teacher, and students.

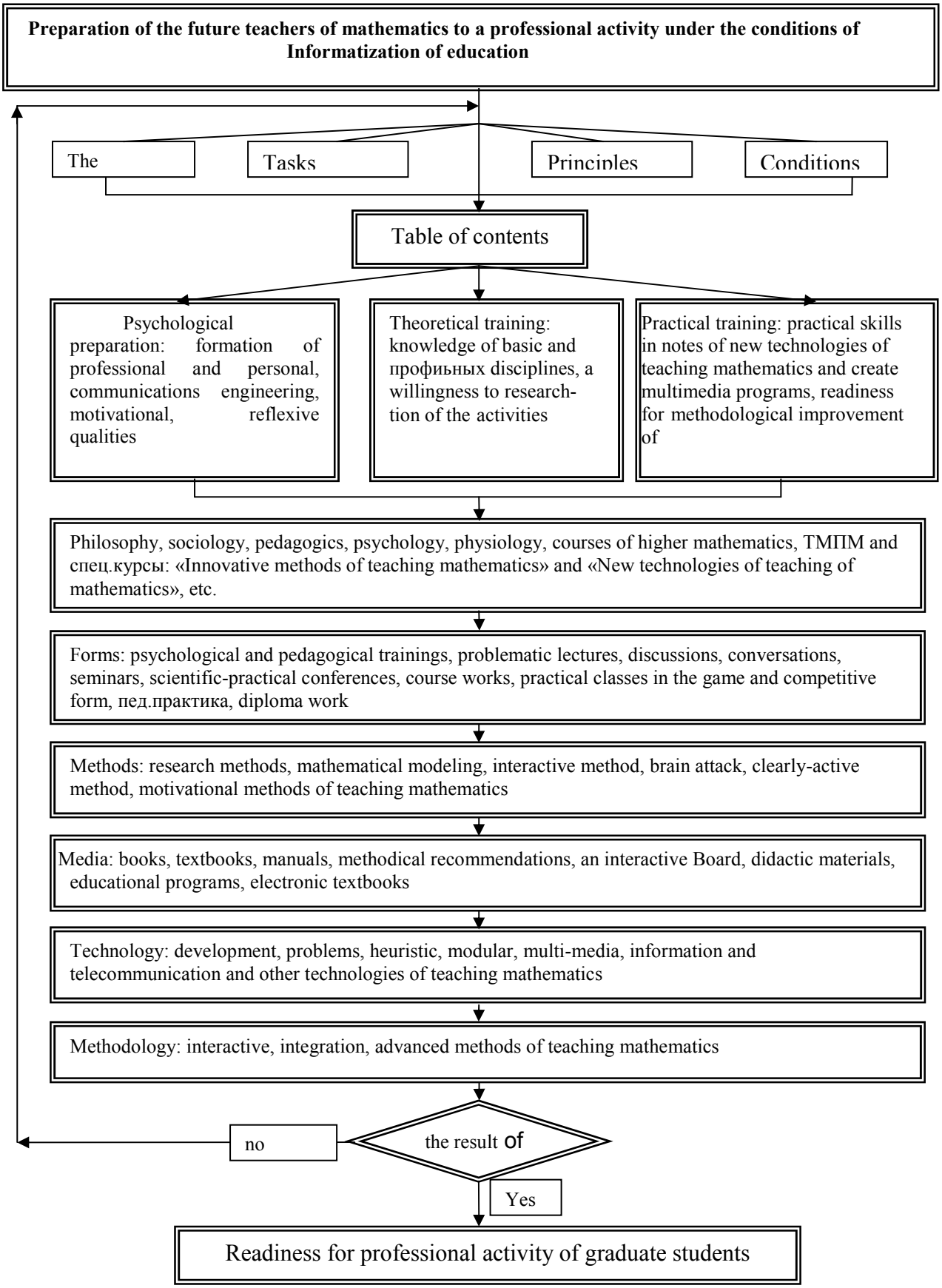


Fig. 1. **Model of training the future teachers of mathematics to a professional activity under the conditions of informatization of education.**

All it is, accordingly, entail conversion of the main components of the educational process: the changing nature of the joint activity of its subjects; the ratio of the didactic functions, implemented in the system «teacher-computer - student»; the complex of programs and technologies of teaching mathematical disciplines; modified methods and forms of training.

Informatization of society is at the present time objectively happening process. Its feature is that one of the main activities of the members of the society are processes related to the collection, storage, processing and translation of information.

In connection with this, one of the leading directions of the process of informatization of a society is the process of informatization of education, which provides education methodical and practical tools for creation and use of information technologies for the purposes of training and education.

One of the main issues of informatization of education is preparation of future teachers of mathematics on the basis of modern requirements of pedagogical science. At the present time at our Department there is some experience in the study of the problems of theoretical and methodical preparation of students of teacher training colleges.

The problem of preparation of the future teachers of mathematics at pedagogical universities of is carried out by means of the following forms, means, methods and technologies.

Forms: psychological and pedagogical trainings, problematic lectures, discussions, conversations, seminars, scientific-practical conferences, course works, practical classes in the game and competitive form, degree work;

Equipment: electronic educational and methodical complexes (EMC), textbooks, manuals, methodical recommendations, an interactive Board, didactic materials, educational programs, electronic textbooks.

Methods: research methods, mathematical modeling, interactive method, «brainstorming», clearly-active method, motivational methods of teaching mathematics.

Technology: development, problems, heuristic, modular, multimedia, telecommunications and information technologies of teaching mathematics.

Use of the possibilities of multimedia technology for presentation of audio-visual information for educational purposes and developed on the basis of multimedia courses significantly extend the capabilities of educational material presentation including through the inclusion of animation, sound and video. In addition to the new forms of presenting the information programs, developed on the basis of multimedia technology, have another valuable from the methodological point of view of quality - of interactivity, which allows more completely realize the learning effect of these training tools.

Wide potential possibilities of multimedia technology were put in front of pedagogical science problem on two interrelated areas: methodological aspects of the application of multimedia technology for mathematics and methodological aspects of creation of multimedia programs of academic appointments in the system of training the teachers (Portfolio, 2005).

The rapid development of telecommunication networks and based on their information technology has a huge impact on all areas of the education system, both in our country and abroad. The impact of new information technologies in the education system appears in two ways. On the one hand, these technologies allow increase of the efficiency of educational process, scientific research and management of educational institutions.

On the other - the system of education, especially higher, is an active participant in the process of development of information technologies and information resources.

It is important to note that the role of the teacher in the conditions of informatization of education is not only the leading, but also is enhanced even more. This is due to the fact that the teacher carries out its in the new educational environment, which is characterized by the use of modern means of information.

Thus, the content of activity of the teacher of an increasingly creative nature, which requires from him constantly update their knowledge and professional growth. The problem of quality of preparation of teachers to work is particularly acute in the principally new conditions of their professional activities.

Experience of the KazNPU named after Abai

In this connection, the methodical Council of the Institute of magistracy and PhD studies at the KazNPU named after Abai hold training seminars for the staff (PPP) of the Institute:

1) the technology of e-learning;

2) interactive technologies in teaching.

And also the administration of the Institute pays attention to:

- the provision of the disciplines of the departments of the Institute of educational-methodical literature;
- to the material-technical provision of the educational process;- on introduction in educational process of modern information and educational technologies.

In the block of natural science disciplines all the professional educational programs of scientific-pedagogical directions included special courses «New technologies of teaching mathematics», «the Basics of e-learning and virtual innovation», who are currently performing the role of interdisciplinary science, technology, and methods which are integrated in the General-professional and special disciplines of the training of future teachers.

The teaching of mathematics has its own specific characteristics, though the methods of teaching and method of organization of educational material, it would seem, is traditional theoretical information, practical studies, conducting the control measures. This is due to the fact that all branches of mathematics are developing rapidly. That requires constant updating of the educational-methodical ensure that the most optimal for the electronic submission of the educational-methodical and information support. It should also be noted that specific feature of the study of mathematics with the help of information technologies, as the «duality» of its basic element - a computer, which is both the object of study, and at the same time learning tool and instrument of accomplishing the tasks.

Specific peculiarities of studying mathematics with the help of information technologies bring to the forefront the problem of such organization of educational process, in which the student is the subject of the educational process, the individual aspiring to self-realization and self-government of its academic activities, and the teacher becomes the organizer of independent active cognitive activity of students. In these conditions the realization of «subject-subject» of relations communication teachers and undergraduates will become joint creative activities, but not limited to the transfer of knowledge. In higher educational institutions of the primary organizational form, in the framework of which formed the foundations of subject - subject relations, are the lectures. The lecture is a leading, the base element of didactic training cycle: its aim is the formation of an indicative basis for the subsequent learning by the students of educational material.

The peculiarity of lectures in comparison with other organizational forms of the educational process is its activity framework, which causes the need for training undergraduate independent study activities, including the following:

- the implementation of the necessary preparation for the lecture classes;
- mastering the methods of sustained attention and active comprehension of material during the lecture;
- development of their own management systems taking notes of the lectures;- the addition of the abstract after the lecture;
- the absorption of lecture material, self-control and execution of creative independent works on the basis of the lecture material (Kaskatayeva, 2011).

Thus, the lecture is teaching the organizational form of the educational process, which is a very economical way to get the fundamentals of scientific knowledge, is a powerful means of activating the cognitive activity of undergraduates, the means of development of the vision of the problems and abilities of independent identify approaches to their solution, as well as contributes to the development of the masters skills of self-government of its academic activities.

The actual solution of optimal organization of educational process in high school-e, as it seems to us, is the creation and application of multimedia lecture complex at the highest mathematics, which includes not only didactically-informational tool, multimedia lecture Annex, module testing the knowledge of lecture topics, but also the module of self-organization of extracurricular work on the lecture material.

In the beginning of the academic year graduates hear a syllabus, where there are:

- abstracts of lectures;
- a brief description of the seminars and practical classes (plans, tasks to conduct seminars and practical classes, iwst, CDs);
- themes and a brief description of the laboratory and Studio work;
- tasks for self-evaluation and preparation for the examination, including tests, a list of basic and additional literature, including in electronic media;
- the list of Internet resources;
- glossary.

Presenting the lecture material, the teacher focuses on the fact, as undergraduates outlines the training material. The deliberate note-taking includes listening, comprehension, processing and brief record. The lecturer of the following: all the students understand and time to process the academic information. It is very important to the lectures of feedback, which when using the multimedia lecture complexes can be implemented in the form of mini-self-test, which is carried out after each topic of the lecture material. Graduates on the screen is provided 3-5 issues, and every listener, in answering these questions, controls their assimilation to the taught material, comparing their answers with the answers, which opens a lecturer on the expiration of the time allocated to the self-test.

At the end of the lecture the lecturer pays attention to the most difficult questions of the lecture material and recommends educational literature, as well as periodic publications on these topics for extracurricular work with the Ukrainian educational material.

The most important condition of mastering the material you listened to the lectures is extracurricular independent educational activity of students on the deeper comprehension and memorization, which includes several interrelated stages. The first stage of this activity is to read and substantive revision of the outline of the lecture.

On the practical (Seminary) lesson after the technical design of the abstract should be re-reading of the lecture material with a view to a more profound understanding and digestion of the material. At the same time he should

be aloud or silently reciting the key provisions of the lectures, definitions, formulas, conclusions, etc., because multiple read without playback does not provide knowledge.
The periodic repetition of the lecture material is the means of lasting scientific knowledge mastering subject areas and creates preconditions for long-term retention and assimilation of the studied material, as every new treatment to one and the same material opens up new facets of, who had previously escaped from the attention.
During the time of independent work of master students under the guidance of the teacher (IWMSUGT) conducted the study of scientific and educational literature on the themes of the lectures, the expansion and deepening of the knowledge, acquired skills to work in subject areas. This stage it seems to us very important from the point of view of the development of the masters skills of self-government of its academic activities.
In the process of work with the scientific editions develops creative scientific thinking, is forming information culture of the future teachers. But the theoretical study of the scientific and educational literature on special courses of higher mathematics necessarily should be complemented by the implementation of creative independent works. Only in this case the independent educational activity undergraduates will contribute to the assimilation of knowledge of the subject region, development of the ability to work with information: analyze, compare, organize, classify and generalize, ownership of abilities and skills to apply the knowledge in practice.
As practice shows, systematic monitoring of the implementation of independent creative work on the lecture material from the teacher with the traditional organization of the educational process is difficult. It is necessary to involve themselves undergraduates in the process of self-assessment of knowledge and skills obtained in the course of studying of mathematical courses. For the training of undergraduates analysis of their knowledge and training actions we offer them to create a collection of his works on the mathematical rate.
And if the previous stages of the lecturer are controlled by the periodic review of abstracts of lectures, conducting workshops, the interim reports, the control IWMSUGT and independent work of master's degree students is a method of portfolio.
At the present time, this method of evaluation of quality of knowledge of students - the method of the portfolio is of great scientific and practical interest. Analysis of scientific-pedagogical literature shows that this method is mainly found a use in the school and in University practice (Chernilevsky, 2002). For teachers of the higher school of this method is applicable for the training and promotion of a masters the skills of self-government of its training activities, including training lecture.
The method of portfolio is a method of evaluation and self-evaluation of knowledge, which involves the presentation and documentation of their knowledge in the subject areas, as well as taking into account the ability of pupils to the decision of non-trivial tasks and skills for joint work.
The portfolio as a method involves the creation of a «portfolio of student». Problem-based portfolio in mathematics, as an information and control the means of educational activity of master, reflects the dynamics of the educational, labor, intellectual activity; level of skills and skills in the field of information technologies; develops an interest in mathematics, critical attitude to his activity; forming and developing the skills of self-government of its academic activities, and teachers provides optimal variant of the evaluation of various educational results the activity of graduate students: academic, creative, labor, management, and also allows you to evaluate the dynamics of the development of information culture of the author of the portfolio.

Conclusions

This, information and communication technologies (ICT) in education are needed in the preparation of future teachers of mathematics to the professional activity in the conditions of informatization of education, as:
- ICTs can contribute to the solution of some problems of education;
- ICT increase the motivation to learn and play an increasingly important role in society;
- ICTs can contribute to the implementation of a new perspective on learning and teaching.

How to place information and communication technologies in the training programme? There are two main possibilities. First, ICT can be an object, the object of which will be computer literate. Secondly, ICT can be an important aspect of the study related to professional practice. ICT is also a means for learning and teaching.
Future teachers need to teach the correct use of ICT in the learning process. ICT as a means of learning and teaching tool that helps implement innovative methods of training.
The potential for learning using ICT are as follows:
- the creation of real contexts for learning;
- the multiplicity of types of visual presentation;
- encourage the active and reflexive learning;
- increase of the graphic skills;
- increasing the effectiveness the use of time for the solution of the problem.

In the use of ICTs in education need to take into account a number of conditions:
1) the condition of using the potential of ICT in education is the need to restructure the curriculum and a training plan, which should have a greater focus and more depth,
2) And also is necessary in syllabus give thorough job for active work of graduate students and stimulate their thinking.

Thus, the necessity for the use of information and communication technology (ICT) in education and developed a model of training the future teachers of mathematics to a professional activity under the conditions of informatization of education.

References

Bent, B., 2005. Andersen, Katja van den brink Multimedia in education: specialized training course. - M.: «Training-Service», 2005. - 216 p.

Chernilevsky, D.V., 2002. Didactic technologies in the higher school. - M.: Unit - Dana, 2002. – 437p.

Kaskatayeva, B.R., 2009. Development of future Mathematics teachers' methodological competence, Monograph, Almaty, 2009.

Kaskatayeva, B.R., 2011. Methods and technology of teaching Mathematics , Tutorial, Almaty, 2011.

Portfolio in the modern educational field: educational-methodical manual /Under the editorship. N.N. Surtaevoi. - St. Petersburg-Tyumen: TOGIRRO-NMCs, 2005. – 40p.

Respondents´feedback on online learning reflecting individual learning preferences

Ivana Simonova, Petra Poulova

Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic
ivana .simonova@uhk.cz; petra.poulova@uhk.cz

Abstract: *The three-year research project "A flexible model of ICT-supported process of instruction reflecting individual learning styles" has been running at the Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic. The main research objective was to verify whether reflecting individual learner´s preferences within the ICT-supported process of instruction result in improving students´knowledge in comparison to the process reflecting teacher´s style or being self-managed by the learner. The paper provides results of monitoring students´feedback after studying in the online course. After finishing the process of instruction students expressed their opinions, experience and attitudes in the final questionnaire, containing 22 items providing detailed information about the sample group (five items); and seventeen questions dealt with respondents´ experience in studying the course. Despite the learning process was provided in three different versions, most students expressed their satisfaction with the way of instruction they underwent. This result was deeply analyzed and several interpretations provided.*

Keywords: tertiary education, online, learning styles, individualization, questionnaire, feedback.

Introduction

Since 2010 the research project "A flexible model of ICT-supported process of instruction reflecting individual learning styles" has been solved at the Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic. The main objective of the research was to verify whether tailoring the process of instruction in online courses to learner´s individual learning style and preferences results in improving students' knowledge in comparison to the traditional teacher´ style led instruction and self-managed process (Felder, 2010), (Šimonová, Poulová, 2012), (Kostolanyová, 2012). The project was structured into four phases: (1) detecting students´individual learning styles; (2) designing the online course "Library services – Information competence and education" in three versions reflecting (a) students´ individual learning style, (b) teacher´s style of instruction and (c) monitoring the process of instruction managed by each student individually; (3) running the pedagogical experiment to verify or reject the hypothesis that tailoring the process of instruction to student´s preferences results in better knowledge expressed in higher test scores; (4) monitoring students´feedback after the process of instruction.
The questionnaire focused on the phase four of the process.

Research Methodology

Nine topics were included in the learning content of the online course: Basic terminology, Library services, Bibliographic search services, Secondary sources, Electronic sources, Writing professional texts, Bibliographic quotations, Bachelor and diploma theses, Publishing ethics. Study materials for each topic were prepared in several forms: (a) full texts providing detailed information; (b) short texts structured for the distance form of education, where the structure of the material aspires to simulating the traditional, face-to-face process of instruction; (c) PowerPoint presentations; (d) animations; (e) video-recorded lectures; (f) links to additional sources and other types of study materials. Each chapter included not only the learning content but also examples, practical applications and individual activities in such forms appropriate to the online course concept (version) described below. The sample group consisted from nearly 400 respondents from the Faculty of Informatics, University of Hradec Kralove.
The course was designed in three versions reflecting:

- *Respondent´s style of learning.* Students in the experimental group 1 (group LCI) were offered such study materials, exercises, assignments, ways of communication and other activities which suit their individual learning styles detected by the Learning Combination Inventory (LCI). The preferred order of materials was made electronically by an application which automatically generates the "offer", i.e. it provides each student with types of materials appropriate to his/her learning style.
- *Content general approach.* Students in experimental group 2 had access to all types of materials (CG – content general) and the process of selection the appropriate type was the matter of their individual decision (group CG).

- *Teachers style of instruction.* Students in the control group (group K) studied under traditional conditions, when their course was designed according to the teacher´s style of instruction which they were made to accept.

Before the process of instruction started, individual learning style of each respondent was detected. The Learning Combination Inventory (LCI) by C. A. Johnston (1996) was used. It includes 28 multiple-choice questions and three open-answer ones. The results is defined by a "pattern" of individual style of learning which presents the combination of four approaches to processing information - the Sequential, Precise, Technical and Confluent Processor.

The learning style pattern structure of all research groups did not differ significantly. Final results are displayed in fig.1 and table 1. The first feature which is apparent when comparing the four figures is that hardly any differences can be seen in comparison the three groups. So, all groups were considered equal and no statistic verifications were required.

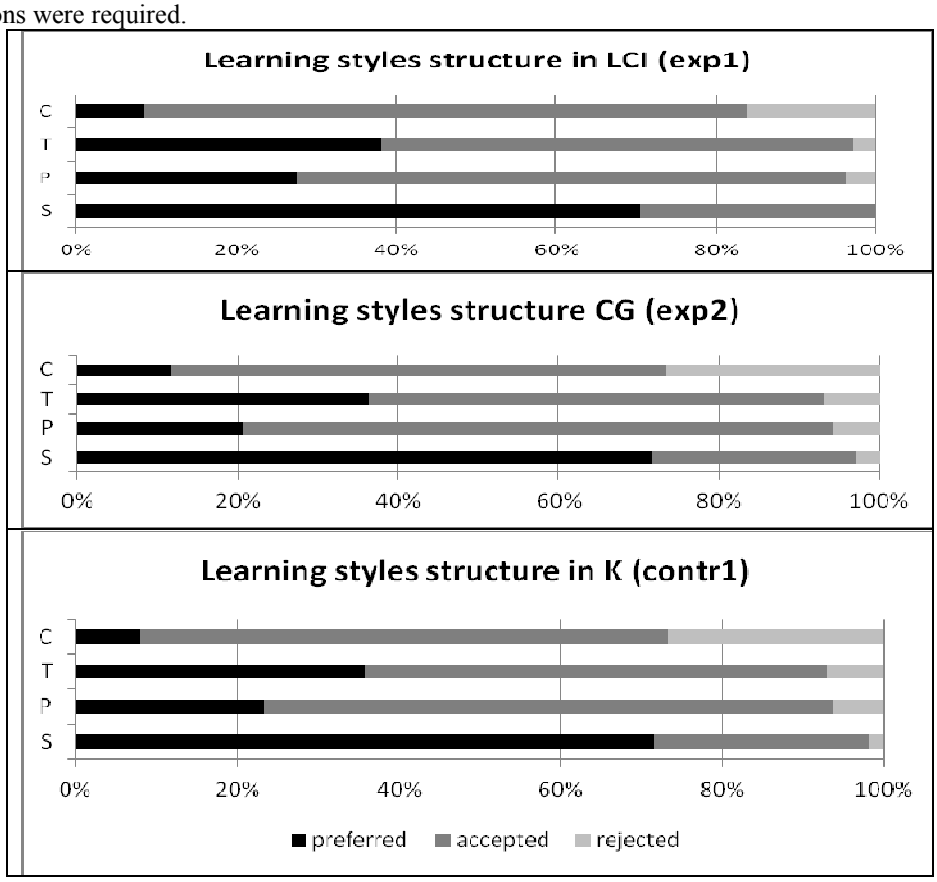


Fig. 1. **Learning styles structures in the LCI, CG and K groups.**

Table1

Learning styles structures in the LCI, CG and K groups

Group/n		Sequential	Precise	Technical	Confluent
LCI 108	preferred	75	29	40	9
	accepted	33	75	63	82
	rejected	0	4	3	17
CG 103	preferred	76	23	39	14
	accepted	26	76	59	64
	rejected	3	6	7	27
K 113	preferred	80	26	41	9
	accepted	31	80	64	74
	rejected	2	7	8	30

The experiment having been closed, students expressed their opinions, experience and attitudes in the final questionnaire, both from the technological and didactic point of view. The questionnaire included 22 items as follows:

- detailed information about the sample group was provided in five items;

- seventeen questions dealt with respondents´ experience in studying the course – seven statements were evaluated on the four-level Likert scale, four items used the four-level classification (1-best, 4-worst) and six ones were the open-answer questions.

Research Results

Data collected in the first part of the questionnaire described the main characteristics of the research sample. The sample group (400 respondents in three groups, 230 – 235 each) consisted from 60 – 63 % of men in all three courses. Respondents were from 20 – 50 years old, approximately 80 % in the 20 – 24 year-old group. Respondents mainly graduated from secondary professional schools (62 – 67 %), followed by grammar school graduates (29 – 45 %). Most of respondents (60 – 65 %) did not have any previous experience in studying online courses, approximately 20 % of them had studied one course and 5 % were experienced online learners having passed four or more courses.

The second part of the questionnaire provided data on students´ satisfaction with the process of instruction in online courses.

First, respondents expressed their satisfaction with clearness of instruction on how to study in the course and whether the organization/re-organization of study materials on the home page reflected/non-reflected their preferences. More than 80 % of respondents considered the instructions clear and 95 % and more respondents expressed their full and partial satisfaction with the way how study materials were re-/organized on their personal home page of the course, i.e. various types of study materials were presented in such order which reflected students´preferences.

Another problem considered by respondents was the length of study relating to the learning content and amount of events scheduled in their diaries. Most respondents (64 – 68 %) did not have problems to master the learning content in the given period but one third of them (32 – 35 %) would appreciate slightly more time. On the other hand, 41 – 52 % found the three-week period for studying the course in their schedule without any problems, followed by another group of 37 – 46 % of respondents who might have had some slight problems and expressed partial satisfaction.

Following two items dealt with evaluation of various types of study materials from the point of respondent´s preferences. All materials were categorized in seven main groups described above (i.e. fulltexts providing detailed information; short texts structured for the distance form of education, PowerPoint presentations; animations; video-recorded lectures; links to additional sources and other types of study materials). Four of them (fulltexts, texts for the distance education, PowerPoint presentations and animations) were evaluated by respondents from the point of their appropriateness to the individual learning style (from 4 – fully matched to 1 – completely mismatched to my learning style), when each value can be used once only.

As mentioned above all groups were stated identical from the point of learning patterns structure (Fig. 1, Table 1). The results of the experimental group 1 (LCI), where the study materials were re-organized by the plug-in so that the process of instruction reflected individual learning styles, show rather strong preference of full text study materials (1 – 39 %, 2 – 22 %) and texts structured for distance learning (i.e. where the applied methods, tools and style of writing simulate the real process of instruction, 27 %, 37 %), while approximately half of respondents (20 %, 33 %) most appreciated presentations and animations are least preferred type of study materials (23 %, 21 %).

Respondents of the experimental group 2 (CG), who were provided all types of study materials and no re-structuralization was made by the plug-in, also show strong preference to full texts (40 %, 23 %) and texts structured for distance education (27 %, 3 %); presentations show the lowest preference in all three groups (13 %, 31 %), they are even less appreciated that animations (21 %, 17 %).

Results in the control group (K), where the process of instruction reflected the teacher´s style, preferences were nearly identical with the experimental group 1 where the learners´ preferences were reflected, i.e. fulltexts were the most appreciated type of study materials (25 %, 35 %) followed by texts structured for distance education (23 %, 26 %); presentations received the highest rate of appreciation from all groups (24 %, 30 %) and animations are in-between the experimental group 1 and 2, reaching 32%- and 14%-preference, which is higher rate than in the LCI group and lower than in the CG group.

If the collected data are transformed to nominal values (4 points for mark 1 – this type suits best and 1 point for mark 4 – fully does not suit), results clearly show

- very positive appreciation of fulltexts in all three groups followed by partially positive evaluation;
- comparable rate of full and partial appreciation of texts structured for distance learning;
- rather high rate rejecting presentations;
- strong appreciation of full texts in the control group where instruction follows teacher´s style,

as displayed in Fig. 2.

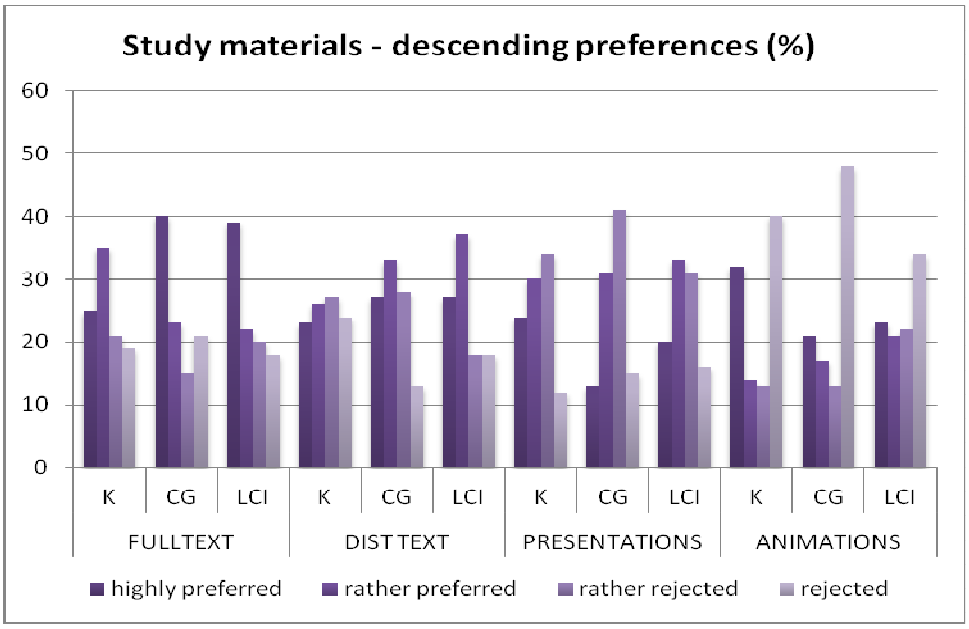
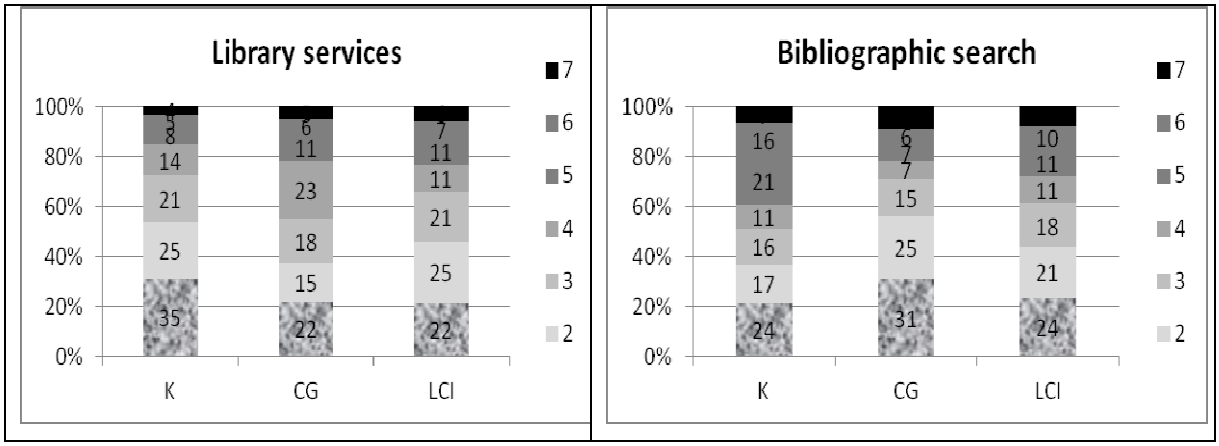


Fig. 2. **Study materials in descending preferences.**

Furthermore, several items collecting students´ feedback on the difficulty of mastering the learning content were also included in the questionnaire. The learning content was structured in nine above mentioned chapters. Fig. 3 displays respondents´opinions on the difficulty of selected topics. The difficulty was evaluated by on the seven-level scale from value 1 – least difficult topic (dark colour) to value 7 – most difficult topic (light colour). While mastering the topics of Library services, Bibliographic search services and work with Secondary sources were considered rather easy than difficult (dark colour – easy to master, light colour – difficult to master), other three topics were evaluated rather difficult (levels 6 and 7) – i.e. Electronic sources, Creating quotations and Professional writing. Two items are surprising within this evaluation: first, ranking work with electronic sources among difficult topics by today´s students who declare modern technologies to play irreplaceable role in their (both private and professional) lives; second, at the same time their evaluation of mastering bibliographic search services rather not difficult.

Creating quotations is considered the most difficult topic (level 7) by 40 % of respondents in the control group (K) reflecting the teacher´s style of instruction and 25 % on level 6), followed by the LCI group (experimental group 1) in which individual learning styles were reflected in the process of instruction (35 %, 25 %) and by the CG group (experimental group 2; 29 %, 16 %).

Professional writing was also regarded as a difficult topic; 26 % and 28 % of respondents in the control group evaluate it very difficult (levels 7 and 6), and similar results appeared in other groups: respondents in the LCI group showed 27 % and 25 %, in the CG group the results were 24 % and 31 %. The complete data are presented in fig. 3.



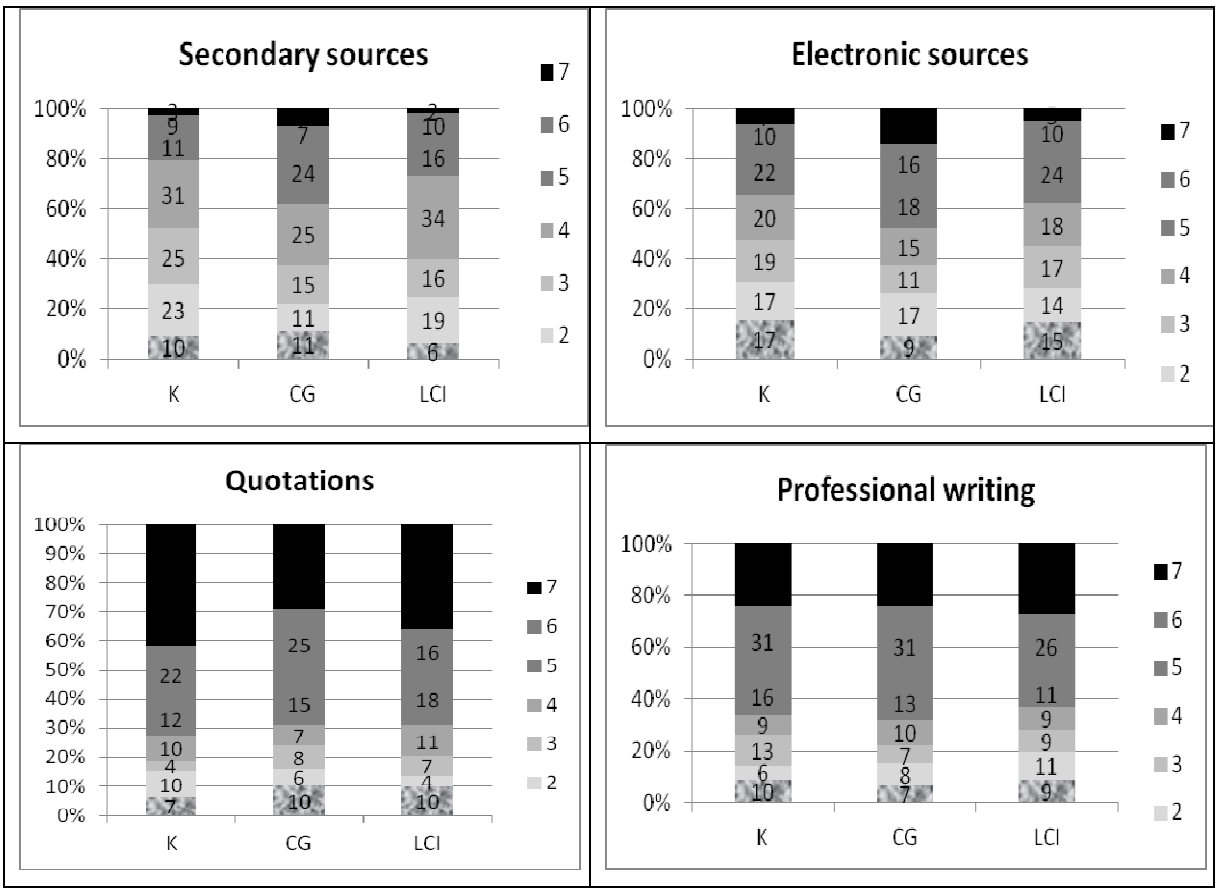


Fig. 3. **Difficulty of selected topics** (1 – least difficult, 7 – most difficult).

If the learning style preferences are not taken into account, results of total difficulty evaluation are displayed in fig. 4. The darker the colour is, the less difficult (i.e. easier) the mastering the topic was for students. Thus it is clearly visible that chapters of Quotations and Professional writing are evaluated the most difficult, similarly to fig.3.

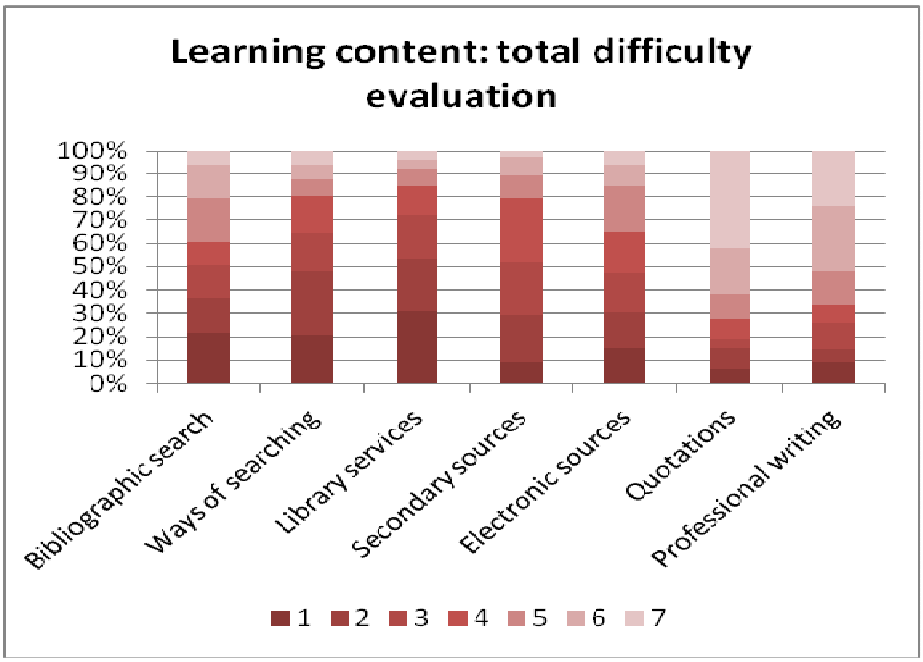


Fig. 4. **Total difficulty evaluation of the learning content** (1 – least difficult, 7 – most difficult).

The course of study was also evaluated from the point of learners´problems, difficulties and limits. Five criteria were set as follows:

- to start studying,
- to keep studying,
- lack of time,
- tiredness,
- problems with technology.

Data were evaluated on the six-level scale from no problems (level 1) to crucial problems (level 6). Results are presented in fig. 5.

Fig. 5. **Selected matters within the process of instruction.**

Half of respondents (48 %) had no or little problems (levels 1 – 3) to make efforts and start studying in the LCI group while slightly fewer ones were detected in CG (46 %) and K (44 %) groups; higher rate was even expressed under the second criterion, i.e. keep studying, when respondents in the K group reached the highest score of 61 %, followed by the LCI group (55 %) and CG group (53 %).

Approximately 20 % in each group suffered from lack of time for studying (61 % in the LCI group, 54 % in the CG group and 52 % in the K group.

What is rather surprising students did not feel so much tired as could be expected when studying in the combined, i.e. part-time form. Fifty-nine per cent of respondents in the LCI group had no or slight problems with tiredness before or within learning, and even the higher scores were reached in other groups - 63 % both in the CG and K group.

Most respondents did not have substantial problems with technology (85 % in the LCI group, 79 % in the CG group and 84 % in the K group).

Research results and discussions

From the above presented results can be clearly seen that respondents – participants in the online course, expressed their positive approach and satisfaction with the course of study. Hardly any crucial problems appeared which could be also caused by the fact, that the respondents were students of the Faculty of Informatics and Management (Applied Informatics and Information Management study programmes, future engineers) and neither the online learning (i.e. ICT-supported instruction), nor the entire learning environment built limits and restrictions to them in the process of forming knowledge.

Unfortunately, neither the research results, nor learners´evaluation proved our expectations that the reflection of individual learning style might be the means which (if applied in the didactic and sensitive manner) could help them substantially within the process of online learning (Honey et al., 2000). This result was surprising for the team because the learning style reflection was understood to be a powerful factor providing strong impact on the process of learning, and statistically significant increase in knowledge of the LCI course participants was expected.

Despite all the possible and real problems the course participants had during the course of study, approximately 80 % would take another course (other courses) within their university study, i.e. they would prefer online learning to traditional face-to-face approach.

There might be several reasons how to interpret the results.

First, neither strong, nor marginal preferences were discovered in patterns within the sample group which could produce statistically significant differences. We agree with e.g. Honey (2010a), Mitchell (2004), Sternberg (2001) saying that not tailoring the process of instruction to learners´ individual preferences results in increase the knowledge but they consider the developing new learning strategies to be more contributive to the learner. Thus the research question is as follows: Is it really worth dealing with learning styles if the pedagogical experiment did not prove any increase in knowledge?

Second, as mentioned in the theoretical part, there exist some researches (and researchers) that reject the theory of learning styles resulting in the individually tailored process of instruction, e.g. Honey and Mumford (2002), Honey (2010b). The proposal might be to work with learners (a) showing very strong preferences in one learning style, and help them develop other strategies and approaches; (b) attract attention and show those who have very weak preferences and are able to study efficiently using any strategy that there exist some approaches and methods which might suit them better, which finally can increase their motivation in learning, make the process more interesting for them, which is not of little importance (Sternberg, 2010).

Third, there could be several other reasons why the expectations and hypotheses were not verified, both on the researchers´and learners´side. In further research activities other approaches running the process of instruction reflecting individual learning styles can be tested, i.e. tutor´s role as a facilitator could be strengthened and emphasized so that learners feel and study in a more friendly environment, being provided wider technical and didactic support, use their experience in online learning developed in this course, and many other measures could be taken. On the learners´side the skill of independent work and study must be supported and gradually developed, as online learning has become standard not only in the tertiary education but particularly in lifelong learning (Černá and Maněnová, 2010).

One of the project outcomes - the plug-in generating the learning content in adequate order has been designed and can be used as freeware on request but no increase in learners´knowledge was discovered. What has been appreciated is the learners´positive approach to online learning which was expressed by their approach to further learning in online courses. Despite this factor did not belong to the primary or crucial ones, it can be considered a positive side contribution without any hesitation.

Conclusion

Thus it can be concluded that despite the contribution of the learning style theory to the online learning process was not proved within this project, no decrease in learners´ knowledge was discovered in comparison to the traditionally led process of learning which follows teacher´s style of instruction. The above mentioned authors (Gregorc, Mitchell, Honey, Mumford etc.) also described results not verifying the learning styles contribution to the knowledge development and educational objectives within the learning process reflecting individual learning styles; including Felder particularly focussing on engineering education. As mentioned in the first chapter, the time came to deal with didactic aspects of ICT implementation into the process of instruction. Students have not reached higher but the same level of knowledge in online learning, which corresponds to predefined learning objectives; and both teachers and learners have to develop their knowledge and skills towards studying efficiently being supported by modern technologies. This conclusion and recommendation is natural for engineering students and engineering pedagogy, both having close relation to modern technologies.

Acknowledgment

The paper is supported by the Excellence project N. 2216.

References

Černá, M. And Maněnová, M. 2010. Utilization of web Portals and Their Services: A Case Study. *Proceedings of the 9th European Conference on E-Learning*, Vol. 1. Reading : Academic Conferences LTD, 2010, pp. 140-145.

Felder, R.M., 2010. *Are Learning Styles Invalid? (Hint: No!).* Available at: http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS_Validity(On-Course).pdf. 22.08.2012.

Honey, P. et al., 2000. *Attitudes to e-learning: A national survey 2000 undertaken by the Campaign for learning*. Sandford: Southgate Publishers.

Honey, P, 2010a. E-learning: a performance appraisal and some suggestions for improvement. Available at: http://qou.edu/arabic/research/Program/e-learningResearchs/e-learningPerformance.pdf. 11.10. 2012.

Honey, P., Mumford, A., 2002. *Using your learning styles*. Maidenhead: Peter Honey Publications.

Honey, P., 2010b. Learning styles - the key to personalised e-learning. Available at: www.bbmatters.net/bb_matters./Learning%20styles_peter%20honey.pdf. 13.01.2012.

Johnston, C.A., 1996. *Unlocking the will to learn.* Thousand Oaks: Corwin Press, Inc.

Kostolányová, K., 2012. *Teorie adaptivního e-learningu*. Ostrava: Ostravská univerzita.

Mitchell, D.P. et al., 2004. *Learning style: a critical analysis of the concept and its assessment*. London: Kogan Page.

Sternberg, R. D., Zhang, L.F., 2001. *Perspectives of thinking, learning and cognitive styles.* Mahwah: Lawrence Erlbaum.

Šimonová, I., Poulová, P. 2012. *Learning style reflection within tertiary e-education.* Hradec Králové : WAMAK.

Sternberg, R., 2010. Mental Self Government. Available at: http://web.cortland.edu/andersmd/learning/Mental%20Self%20Government.htm. 22.08.2012

Seniors' perspectives on the learning and using ICT: findings within the project AWAKE

Anna Vintere[1], Evija Kopeika[2]

[1]*Department of Mathematics, Latvia University of Agriculture, Liela street 2, Jelgava, Latvia, LV-3001*
[2]*Faculty of Economics and Management, University of Latvia, Aspazijas Boulevard 5 - 332, Riga, LV -1050*
Anna.Vintere@llu.lv, evija.Kopeika@llu.lv

Abstract: *To identify educational needs of people over 50 and improve teaching and developing the offer of education for people over the age of 50, in the framework of Grundtvig project AWAKE (Aging With Active Knowledge and Experience) the survey was carried out in project's partners' countries: Poland, Italy, Romania, Lithuania and Latvia. The questionnaire was divided into four diagnostic blocks: Current situation, Needs and expectations, Possibilities and barriers and Volunteering. Each part contained the questions on the use of ICT (information and communication technologies).*
Active aging includes various areas of human life: participation in the labor market; the household-related activities, including homework and other care; active participation in community life, including volunteering and active use of leisure time for hobbies, sports, travel, creative activities. Of course, a better use of the information and communication technologies can provide significantly greater benefit to the public, increase the individual's role in civic democracy and successful development. Therefore, this paper gives short overview of the project and contains evaluation of the use ICT in education of 50+ people in partners' countries as well as outlines the tips for Europe Union institutions.

Keywords: e-skills, e-inclusion, e-learning, ICT, seniors, training.

Introduction

During the past decade ICT (Information and communication technologies) has become available for the general population. However a gap remains between users and non-users due to several reasons: from missing infrastructure or access, to missing incentive to use ICTs, to lack of computer literacy or skills necessary to take in the part in the information society. Despite increasing levels of ICT usage in all sections of society, the digital divide is still creating a big gap in European countries (EC, 2006). ICT services are present in the everyday life of people and can play an important role in the improvement of everyday life issues such as health, communication, independent living, social contacts etc.
European countries are now facing similar problems connected with the ageing of societies. The number of older people in the population increases fast so it means that one should make sure, that senior citizens are strong and actively participating in social life group. Aging includes various areas of human life: participation in the labor market; the household-related activities, including homework and other care; active participation in community life, including volunteering and active use of leisure time for hobbies, sports, travel, creative activities. Of course, a better use of the information and communication technologies can provide significantly greater benefit to the public, increase the individual's role in civic democracy and successful development. It is very important to contribute to the inclusion of the people aged 50+ so that they can benefit from ICT services according to their specific needs.
In the face of the ageing process of many societies it's vital to take special care of the educational needs of those people, who together with the end of working career are forced to search other developmental opportunities. It's known that people are getting older in the moment when they stop developing. European project A.W.A.K.E. (Aging With Active Knowledge and Experience) is finding the ways how to effectively enable the education and development of people who are above 50 years old. Project is an aiming an exchange of experience and good practices between the partners to promote the mobility, the activation and lifelong learning idea among the over fifty people. With funding from Lifelong Learning Programme - Grundtvig Learning Partnership, the project is delivered by partners from Italy (IT), Latvia (LV), Lithuania (LT), and Romania (RO) and is led by CSI in Poland (PL).
To identify educational needs of people over 50 and improve teaching and developing the offer of education for people over the age of 50, in the framework of project the survey was carried out in project's partners' countries. In this paper given the evaluation of the use ICT in education of 50+ people in partners' countries as well as outlined the tips for Europe Union institutions.

Materials and methods

The questionnaire developed based on the people 50 + educational purposes: living a happy and fulfilled life, which can be divided into three directions: self-development (personal, hobbies), professional development

based on an individual's career development and requirements of labor market, and inclusion in the labor market - preretirement age (the need to acquire new skills) and unemployment (requalification, need to acquire new specialty). All three groups of people who need to be educated require language knowledge, ICT and communication skills. These skills can be obtained in several ways: some people age 50+ continue to study in higher education institutions, attend courses (seminars, conferences etc.), self-taught or joining interest groups (intramural or virtual). It was also taken into account the three possible forms of education:

1. Formal education - a system that includes primary, secondary and higher education levels, the programs is certified by a state recognized educational or professional qualification;
2. Non-formal education - organized outside the formal education and to demand adequate educational activity (certificate);
3. In-formal - self-taught, everyday learning (family, workplace etc.).

The questionnaire divided into four parts – diagnostic blocks: Current situation, Needs and expectations, Possibilities and barriers, Volunteering. Rather each part (except for volunteer work) contained the questions on the use of ICT (Table 1).

Table 1

The instrument of the survey

Diagnostic Block	**Content of Statements**	**Items on the use of ICT**
Characteristics of the respondent - gender, age, place of residence, education, employment, mobility etc.		
I. Current situation (N=3)	Presented statements about the involvement in an educational program / course and how much time spent on studying and dedicating the self-development.	1. Are you currently involved in any educational program / courses on ICT? 2. To acquire new skills or knowledge, I'm a member of specialized social networks in Internet
II. Needs and expectations (N=13)	Presented statements cover the meanings, which are attributed to the learning of people 50+: what they want to learn, why want to learn - motivation, meaningfulness, expectations from learning, what teaching methods preferred, favourite ways of spending free time etc.	1. What would you like to learn*? (How to use computer and the internet)* 2. Which method of learning do you prefer? *(Internet; prefer learning face to face with a teacher, in a group or by himself.)* 3. What's your favourite way of spending your free time? *(Netsurfing)*
III. Possibilities and barriers (N=5)	Statements cover two aspects of people 50+ educational possibilities: are there provided any educational offer in the respondents' place of residence and it quality as well as barriers in participation in education.	Do you see any barriers in your participation in education in your place of living? (*No good connection; Lack of information)*
IV. Volunteering (N=5)	Presented statements about the desire to be a volunteer and in what kind of volunteering interested. Statements cover several potential values of volunteering as well.	-

Four types of the questions are used in the questionnaire form. The questions designated to investigate the attitudes and opinions of the respondents (diagnostic block 1-4): the respondents have to mark several items expressing the person's attitude to the formulated statement. Other group of questions is presented using the modified Likert Scale of 5 scores (diagnostic blocks 2) with the statement are formulated: Strongly disagree, Disagree, Agree, Strongly agree and Likert Scale of 3 scores (diagnostic blocks 3-4): statements and three optional answers expressing the level of a person's experience with the statement are formulated: Yes, No, Don't know. Three open questions are given. Their purpose is to reveal the opinion of respondents of educational needs of people 50+. Four questions with short answers Yes, No.

The sample was selected in the following way:

1. By place of residence: rural / small city / medium town / city - 7-8 respondents in each place;
2. By gender: at least 15 women and 15 men;
3. By age structure: (51-60 - 10 respondents, 61-70 - 10 respondents, 71-80 - 5 respondents, 81 + - 5 respondents).

In total 204 persons participated in the survey. The characteristics of the survey sample are presented in following table.

Table 2

Characteristics of the Sample (N=204)

Quality	Category	Presents
Gender	Female	58
	Male	42
Age	51-60	29
	61-60	50
	71-80	15
	80+	6
Education	higher	38
	post high school	20
	high school	16
	professional	18
	basic	8
Place of residence	rural	16
	small town	19
	medium	33
	city	32
Employment	self-employed	6
	employee	23
	pensioner	66
	unemployed	4
	other	1

Results and discussion

Much of the literature on older people and computers has focused on four main areas: methods of training older people to use computers and computer software; uses computers to improve the quality of life for older people, and attitudes to computers and computing, including perceived barriers and benefits of computer use, and benefits associated with computer use (Richardson at al., 2002).

The current survey results show that 14.71% of all respondents currently are involved in learning ICT by attending special educational program or courses. Evaluating the results of respondents' place of residence (country), currently Lithuanian seniors are the most active in acquiring ICT skills (Fig. 1).

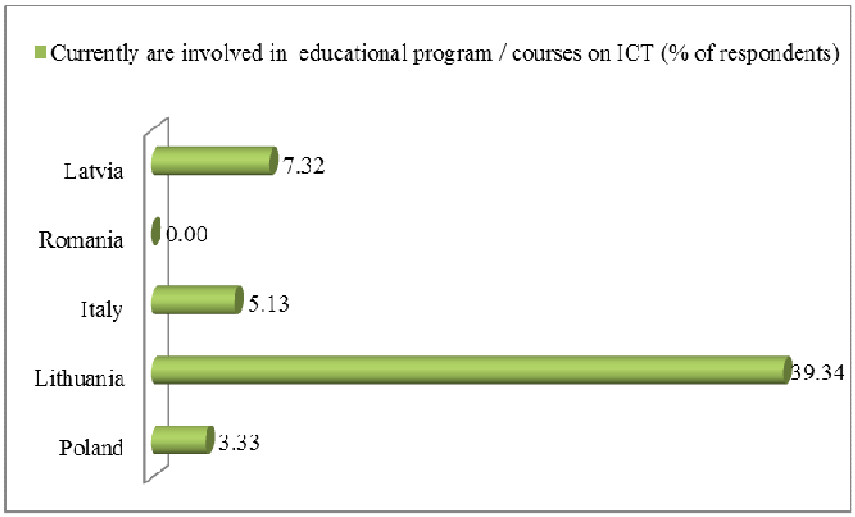


Fig. 1. **Currently are involved in educational program / courses on ICT** (N=204)**.**

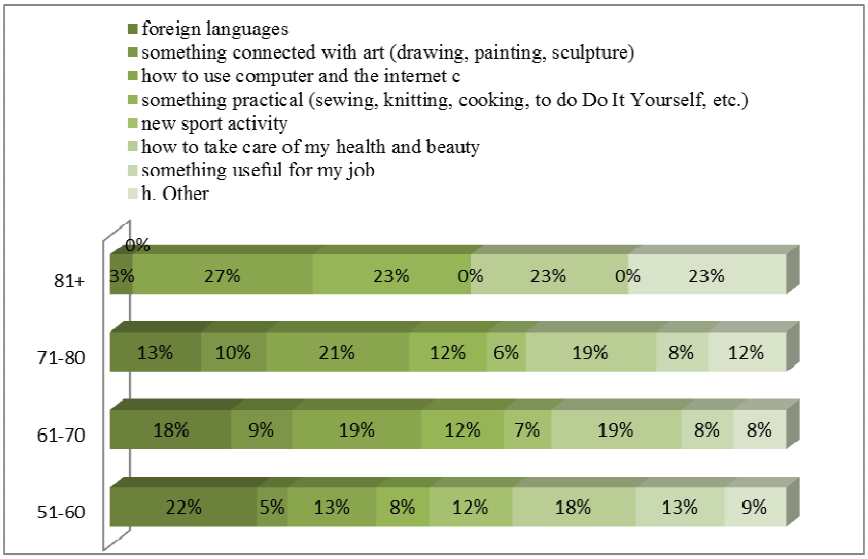


Fig. 2. **People 50+ educational interests** (N=174).

To acquire new skills or knowledge 18.65% of respondents are the members of specialized social networks on the Internet. Latvian seniors are the most active Internet and specialized social networks users for educational purposes (24.39%). Less active are the Lithuanians (6.86%) and Italians (5.13%).In turn, this type of training is not popular among seniors in Poland and Romania.

Statements of the second diagnostic block cover the meanings, which are attributed to the learning of people 50+: what they want to learn, what teaching methods preferred, favorite ways of spending free time etc. Seniors educational interests are given in Fig.2. Results show that greatest interest are the on how to use computer and the internet and how to take care of health and beauty.

Polish seniors have the greatest interest on ICT among the countries (Fig. 3). Less than one-third of the Latvian respondents want to learn how to use computer and the internet. The lowest interest has Romania seniors (only 27.27%).

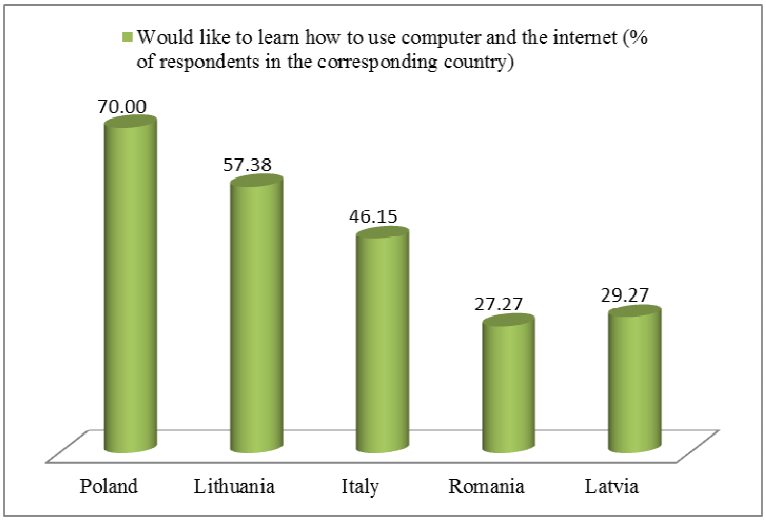


Fig. 3. **Interest to learn how to use computer and the internet** (N=204).

Analyzing the results by place of residence (country), In Lithuania and Italy the greatest interest to computers and Internet have people 50+ in medium towns (accordingly 42% and 20%), but Latvians – in small towns (38% of respondents) (Fig. 4). Romanian and Italian seniors in rural area do not have interest on issues related to ICT.

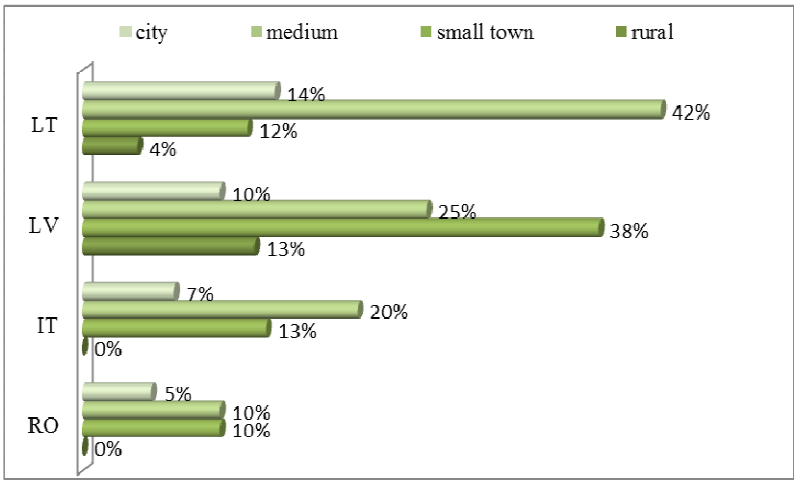


Fig. 4. **Interest to learning how to use computer and the internet by place of residence** (N=174).

Information on interest to learning how to use computer and the internet by educational level is given in Fig. 5. In Latvia the greatest interest have people 50+ with professional education and post high school – both 10%. In turn, In Lithuania the greatest interest has seniors with higher education (43%), in Italy – post high school (19%). The smallest interest on ICT has Latvian seniors with basic education (0%) and high school (2%).

Analyzing the results on ICT as the first priority of educational needs by age structure it is seem that Lithuanian, Romanian and Latvians seniors aged 51-60 and 61-70 have quite similar distribution of ICT educational needs as first priority. In Italy the greatest interest (33%) on ICT is for seniors aged 71-80, but in Lithuania – seniors aged 81+.

Statements of the second diagnostic block cover also what teaching methods preferred by people aged 50+. Based on the survey results there are analyzed Internet as a method of learning by age structure (Fig. 6.), employment (Fig. 7) and gender.

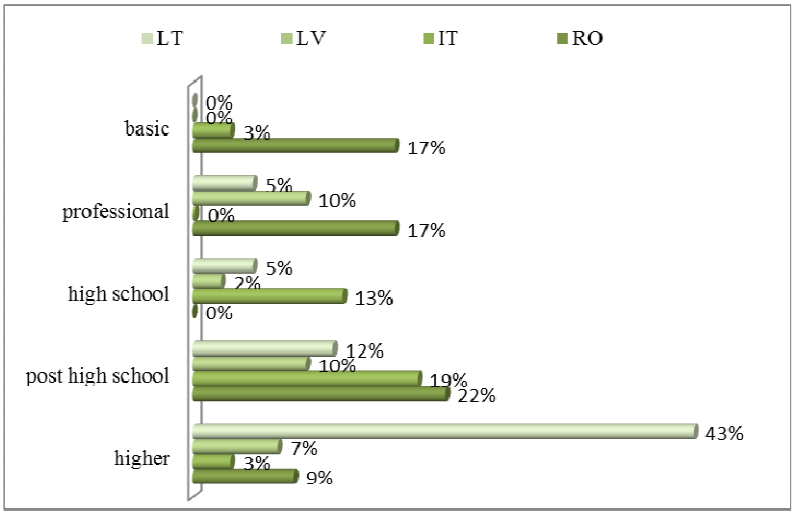


Fig. 5. **Interest to learning how to use computer and the internet by education level** (N=174).

The results show that the use of the Internet for educational purposes is the most popular in Romania, especially for people aged 61-70 (50%) and 71-80 (33%). Internet is the most used for training purposes by Italians aged 61-70 (50%). In Lithuania - distribution of the first three age groups is similar (accordingly 29%, 17% and 27% of respondents), but Latvian seniors' interest to learning through the internet is the lowest among the partner countries.

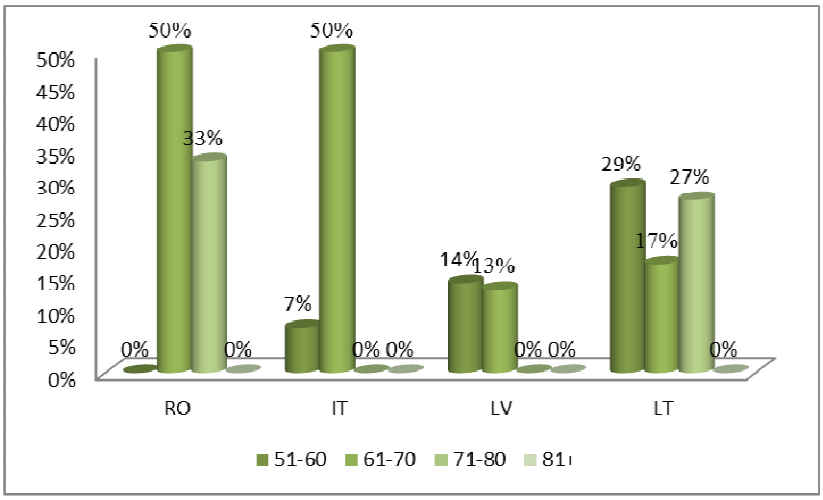


Fig. 6. **Internet as a method of learning by age structure** (N=174).

Latvian self-employed seniors would most like to use the Internet to learn something new (33%), but in Lithuania – unemployed seniors (44%). In Italy distribution among all types of employment are similar (except self-employed seniors), but in Romania the great interest have directly pensioners (10% of all respondents) (Fig.7).

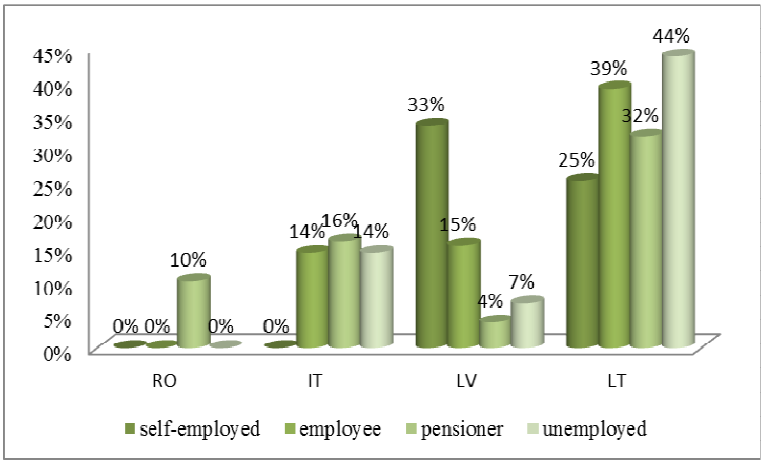


Fig. 7. **Internet as a method of learning by employment** (N=174).

Analyzing the information about internet as a method of learning by gender should be decided that Latvian and Lithuania seniors' interest on the Internet as a teaching method is greater in men than in women (accordingly 17% and 70%), but women – in Latvia 8% and in Lithuania 50%. In Romania and Italy, the situation is reversed - women are more interested in (accordingly 40% and 12%). In turn, to learn how to use computer and the internet by place of residence, distribution by gender (male - female) is the following: Latvia - accordingly 28% and 7%, Lithuania - 20% and 17%, Romania - 11% and 17%, Italy - 8% and 17%.

Several EU policies set the e-inclusion as the need to promote "active" living. In the EU, policies emanating from DG Employment and Social Affairs address active inclusion (linked to the labour market and better access to promote the integration of the most disadvantaged people), decent housing and homelessness, inclusion of vulnerable groups (including isolated older people etc.). The main goal for e-inclusion is improvement ICT access for people with disabilities and senior citizens (E.Mordini at.al., 2009).

Statements of the second diagnostic block cover the favorite ways of spending free time etc. Results show that netsurfing as the favorite way of spending free time recognized 13% of all questioned men and 8% - of women aged 50+. But the distribution by the age structure is as follows: 12% of people aged 51-60, 10% - aged 61-70 - 11% aged 71-80 and 4% of people aged 81 +. It should be decided that netsurfing is not popular among the married people. In Latvia divorced seniors spend the most time on the Internet (33%), in Italy – single (60%), in Romania – married seniors (5%), but in Lithuania – widowers (58%).

Increase accessibility to ICT equipment including assistive technologies to all senior citizens. It is important to recognize that access to ICT and the Internet is by no means universal and lack of ICT infrastructure is a significant barrier in many European countries. Information about access to public Internet provision should be

integrated in information and training programmes for senior citizens (EC, 2006). Statements of third part cover two aspects of people 50+ educational possibilities: are there provided any educational offer in the respondents' place of residence and it quality as well as barriers in participation in education. The results show that respondents have barriers in participation in education in their place of living and one of them - no good connection (Fig. 8). In Romania the quality of connection is almost only barrier to participate in education, but in Poland connection problem cause the least barriers to participation in education.

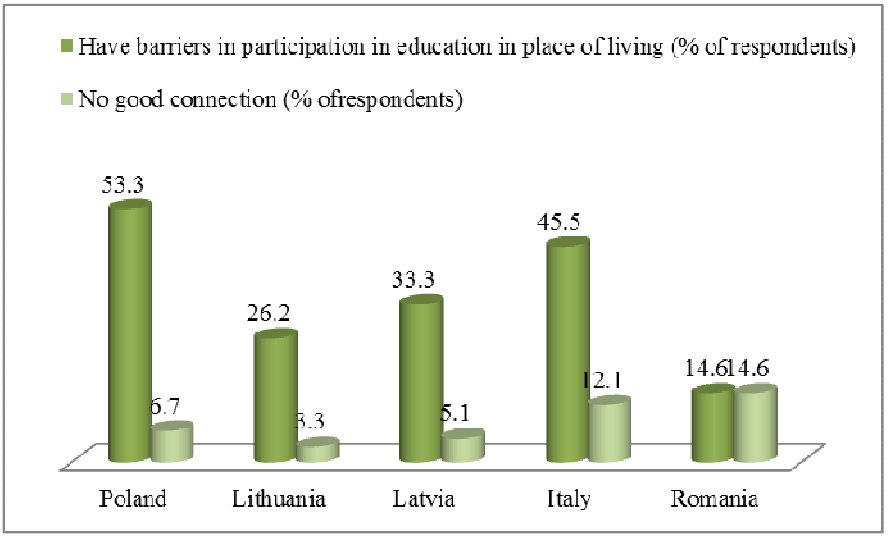


Fig. 8. **Barriers in participation in education in place of living** (N=204).

Conclusion

1. The quality of life literature has focused on lonely, isolated older people who live alone, investigating ways in which technology can assist them to be independent, and ways in which computers can be used to break down their isolation, allowing them to be re-integrated into society (Richardson et al., 2002). Results on the netsurfing as the favourite way of spending free time show that Internet can provide isolated people aged 50+ with increased opportunities for social contacts and can contribute to decreasing loneliness.
2. The questionnaire contains statements on several learning issues: Internet as a method of learning; what teaching methods are preferred - learning face to face with a teacher, in a group or by himself; how old should be a teacher etc. Based on these results could be outlined some tips for the European Union institutions:
   - Use a participatory approach, that involves learners in the process;
   - Use an action-oriented approach – learning by doing; the use of project-based or problem-based activities;
   - In order to motivate seniors to acquire e-skills should be organized short courses that directly related to their daily lives. Seniors will be more motivated to use computers or Internet when they produce visible benefits to them and meet their actual needs;
   - Promote an environment of informal learning for seniors what is not planned in advance with strict program;
   - Use intergenerational cooperation (Junior to senior). However several studies show that sometimes it is unfriendly methods, particularly when being taught by young people who deliver material too quickly;
   - Use the knowledge and experience sharing and transfer method (seniors who have acquired the e-skills train other seniors). Seniors are tended to help each other and in training groups with varied ICT skills it has been observed that more ICT competent learners support those with less skills or slower learning pace (EC, 2006).
3. Measures to promote ICT skills can be: regular (E-Skills Week in the framework of European Get Online Week, courses, consultations, clubs etc.) and various actions organized by Internet Providers, the state / local government bodies or the several interest groups. In the framework of the project are collected best measures examples from all partner countries. Based on literature review and on the Latvian experience of the seniors computer fun clubs are recommended for seniors who want to improve their skills in computer use, to be in the middle of a new appointment and interesting events. The aim of such measures is to interest the wider community on e-skills acquisition and application possibilities, to inform where and how to use existing e-skills or improve them as well as to learn about the e-services and how to obtain it more convenient use.

4. People aged 50+ in Latvia, Lithuania, Poland and Romania is the generation that was educated and started the development of their carriers in „communism" i.e. in conditions of socialism. One can name common characteristics: lack of initiative; think that everything must be provided by state or something else; hold the opinion that it is good to work in the same organization for a long time; not ready for changes etc. Thus providing education possibilities for this generation the primary condition is the enabling environment where is expressed praise very often: praise and praise again!
5. It shall be mentioned that the survey done in the framework of the project AWAKE is a case study which highlights only main the recent trends of the educational needs however it does not foresee any situation at all. The study can become a knowledge background for the development of educational products for people aged 50+ or improvement teaching methods as well as for creation e-learning systems for seniors.

Acknowledgements

The survey was done in the framework of Grundtvig partnership project No. 2011-1-PL1-GRU06-19982 4 "Aging with Active Knowledge and Experience (AWAKE)". The paper's authors are grateful to Centre for Senior Citizen Initiatives (Poland), Centro Servizi Per Il Volontariato V.S.S.P (Italy), Lithuania Adult Education Centres Leaders Association and Asociata Educatia de a Lungul Intregii Vieti (Romania) for the survey conducting and results collection.

References

EC, 2006. Project "Digital Engagement of Senior Citizens". ICT Pedagogic Strategy for engaging Senior Citizens in the Information Society. Available at: http://api.ning.com/.

Mordini, E., Wright, D., Wadhwa, K., De Hert, P., Mantovani, E., Thestrup, J., Van Steendam, G., D'Amico, A. and Vater, I., 2009. Senior citizens and the ethics of e-inclusion. Springer Science+Business Media B.V. 2009. Available at: www.in-jet.dk.

Richardson, M., E.Zorn, T. and Weaver, K., 2002. Seniors' Perspectives on the Barriers, Benefits and Negatives Consequences of Larning and Using Computers. Department of Management School, University of Waikato, Hamilton, New Zealand. Available at: http://ils.indiana.edu/faculty/.

WEB TECHNOLOGIES

Applying Information and Communication Technology To Economic Network Development

Valdis Krebs

Institute of Mathematics and Computer Science, University of Latvia, Raina bulvaris 29, Riga, LV-1459, Latvia
valdis.krebs@lumii.lv

Abstract: *A rural economic development organization needed to build and develop local economic networks in order to transform a local economy from mining & manufacturing to food & tourism. The EDO adopted a two-part strategy: 1) Know the Net — map and analyze the current human network 2) Knit the Net — improve the existing network to increase productivity and learning. Information was gathered on current social and economic actors in the region. A user-friendly software provided network maps and metrics that allowed the economic development organization and their funders to track progress of human network development via information flows, knowledge exchanges, mentoring/advice relationships and business partnerships. The paper will illustrate the process of economic network development using maps and metrics created by the software.*

Keywords: human network analysis, data visualization, network development, resilience, adaptability.

Introduction

Business happens via networks. Whether you have a family business making and selling maple syrup in Canada, or you offer web services over the Internet from your loft in New York City, or you manage a small manufacturing company in Eastern Europe, or you run an international professional services firm in Hong Kong, you are successful because you gain new information, insights and ideas via your personal networks (Krackhardt and Hanson, 1993; Burt, 1995).

Networks are not just for making profits — Non-governmental organizations (NGOs) also benefit from knowing their connections. Networks are important in getting work done (Krebs, 2007), but they are also important in neighborhoods, teams, places of worship, interest groups and other emergent associations. Work is easier when you and your group know where to go for information, advice, and support. Life is easier when your network provides for your needs.

Most of us understand the importance of our human networks. But, other than joining on-line social networks, we do very little to pro-actively build our personal networks. Unfortunately, most on-line social networks do not support network-strengthening activities — they are often like a box where we throw our business cards, clippings, and receipts. Most working-people know who their colleagues are, and how these colleagues may be connected to each other, but they do not know what happens in the organization beyond this very local view. If you work in an organization of more than 50 persons, you are probably unaware of what the others do, and how they connect outside of the organization. When groups in a region, organization, or community are not well connected, internally and externally, they have difficulty with collaboration, learning, and innovation.

Economic Network Development Strategy

How do we build smart, adaptable, learning networks? We start with a map of the current actual situation — the "as is" picture.

- Who is connected and who is not?
- Which groups have formed and how are they connected?
- Who is in key structural roles in the network?

Human networks are complex — people have multiple relationships with multiple others, both near and far. Two people can be workmates, friends, and advisors to each other. Others may just be colleagues or friends, but not both. Still others may not mix work and friendship and therefore they keep their networks separate – non-overlapping.

Some people have many friends, relatives, or colleagues, while others have few. Each person is embedded in a unique location of many networks. They may be on the edge of some networks, while in the thick of things in other networks. Some are in distributed, wide-open networks, while others gravitate towards dense closed networks. They may connect multiple networks, or they may prefer the solitude of one or two networks.

New networks often start as small clusters — whether they are a start-up business, a new NGO, or a community organizing to take some action. Small groups, of often less than five people, get together around a common goal or problem. As they move forward, they start looking for others who can help them — either join together, or provide advice, support or financing. They may look to join with others who are like them. Below is a map of a pre-network — small cluster of people, forming at the same time, around similarities. The small groups are currently not aware of each other.

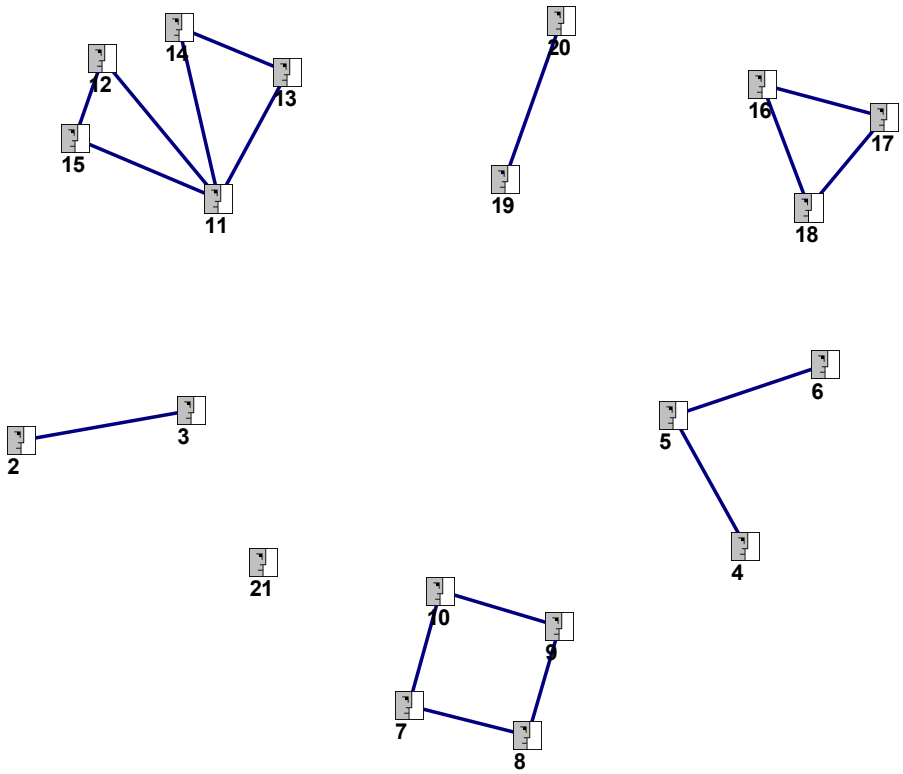


Fig. 1. **Pre-Network: Fragments.**

These could be small community organizations responding to a community need, or they could be start-up businesses focusing on the utilizing the region's resources. They could be employees in a large organization responding to a marketplace opportunity or a new business policy.

Without the proper connections these small teams are unaware what resources, information, and other collaborators are available. They eventually understand the need to make new connections and start searching for NGOs or investors (angels/venture capitalists/government incubators) or a corporate sponsor/expert to help them out. This knowledge source/advisor may connect the small groups together and act as the hub of information and advice as the groups get to know each other. This advisor now forms the most rudimentary network — a hub-and-spoke network, also known as a tree or hierarchy. The hub is the leader, where all spokes connect, and upon whom all spokes currently depend. The people/groups in the spokes do not know each other yet — there is no awareness/trust built between them — therefore they do not share much with each other. Below is a map of the hub-and-spoke network, with the expert/advisor in the role of the hub.

A hub-and-spoke network is very concentrated and vulnerable. With just one node (the hub) holding the structure together, the network is very dependent on the central player. It is a single point of failure. If the hub leaves the network, or fails, the network falls apart into unconnected fragments again. Therefore, most network builders try to get through this single hub phase of network growth quickly. The person or organization at the hub starts weaving the network by making connections amongst the spokes that improve the network.

A map of the hub-and-spoke network is now useful for this next step of network growth. In order to build and strengthen the network, a lead NGO, a network consultant, or a network weaver/builder starts gathering data about all of the network members:

- What are your goals?
- What are your skills?
- What knowledge and skills do you need?
- Who have you worked/collaborated with?
- Who provides you advice/mentoring?
- who do you share ideas with?

This data will be used for two purposes:

1) Add existing nodes and links to the network;
2) Expose opportunities for further links and roles in the network.

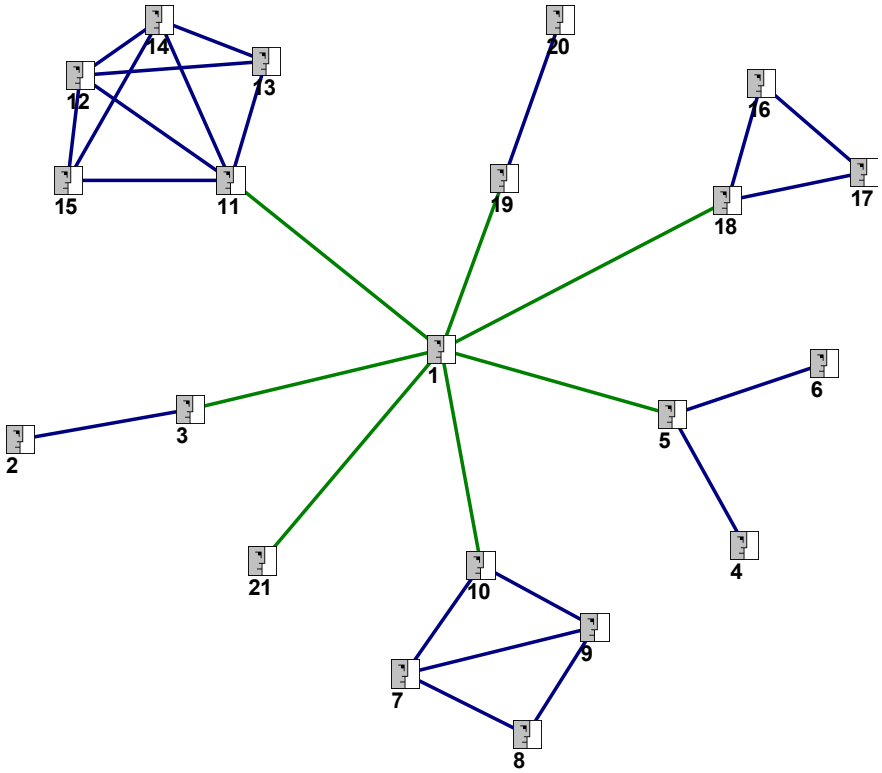


Fig. 2. **Early Network: Hub-and-Spoke.**

After the data is gathered and the "as-is" map is drawn, the network weaver/builder knows the starting conditions for further network development. At this point, if the network is being modeled with the help of information and communication technology (ICT), the group can start looking at various views of the network and can do what-if analysis — what if we add 3 new links in these particular places? Network mapping and network improvement go hand-in-hand. This is the iterative process of "Know the Net" and "Knit the Net". After using the network maps and metrics to build/knit the network over time, the improved network, also known as a “smart network” (Krebs and Holley, 2005), may look like Fig. 3. Notice how some nodes have different roles now than when they started (Fig. 1).

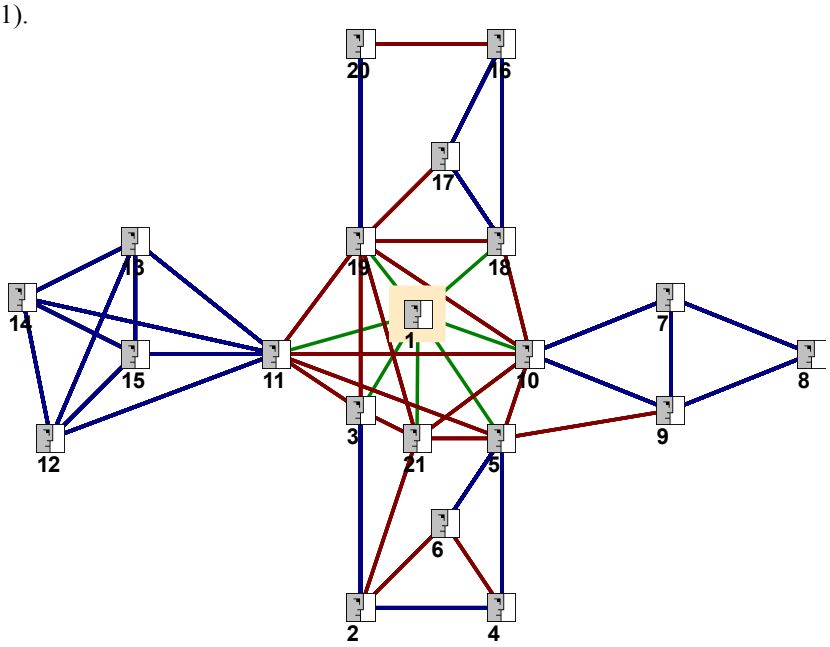


Fig. 3. **Maturing Network: Core-Periphery.**

Network Analysis

Networks can be mapped, and they can also be measured using network metrics. Both the maps and metrics provide useful feedback to the network community and their weavers/builders. Network measures are available

that focus on the individual, the group, and the network as a whole.
Network measures can reveal:

- Who has the most connections?
- Who has the best connections?
- What are the emergent clusters in the network?
- Are clusters/groups in the network interacting with each other?
- What is the overall efficiency of the network?
- Where are opportunities for new connections?

As the network changes and is actively improved, the leaders of the network community will want to track the changes using network metrics. Having an interactive network model, in easy-to-use software, makes this task doable for non-experts. Most professional network analysts have deep training, usually a Ph.D., and expertise in mathematics, data analysis and the social sciences. Not all communities can afford, or can find, such a specialized expert. Therefore, the goal is to make the process of network analysis and network building available to intelligent people who have the need to connect their communities.
The strategy to building a smart/effective network is to create a flexible structure that includes the right people and groups/organizations. The right connections make information/advice easily accessible. The right connections help us navigate the network, and nudge (influence) the network.
A well-developed network has a dense core that is responsible for getting things done — implementation. Dense, redundant ties within the group, help the core group deal with complexity and uncertainty, via trust and learning. The effective network also has a wide-reaching periphery that gathers new information and ideas and makes them available to the core — innovation. Peripheral players in the network connect to their home network and to other networks where critical information, advice and relationships reside. An effective network is successful at both implementation and innovation — both internal and external activity!
The secret to becoming an efficient network is overcoming distance (Wasserman and Faust, 1994). In human systems, distance distorts and delays. The greater the social or contextual distance between two people, the harder it is for them to exchange useful information, advice or have a common understanding. Although, technology like the Internet has almost eliminated the problem of physical distance between people, social and contextual distance remains a problem in human collaboration.
Many have heard the expression of "six degrees of separation" (Watts, 2000) — meaning that any two people in the world are probably connected by five intermediaries (six links). This realization makes it appear that we are all very close to each other — six is a small number when you consider the billions of people on this planet. Yet, six is actually a very large number when it comes to social distance. A Mexican entrepreneur may be six degrees away from a Chinese rice farmer, but they probably don't have much in common in terms of context, goals, language, worldview, etc.
Research in social networks has shown that the important and useful links in human networks are 3 steps or less. Noah Friedkin showed that there is a horizon of observability and influence in networks (Friedkin, 1983). This network horizon is usually around two steps in a network — A knows B and C knows B (but not A), therefore A and C are two steps apart in the network.
Mark Granovetter in his famous study of professionals looking for a job (Granovetter, 1995), found that the chain of job information never included more than 2 intermediaries (a 3 step path) — and these were very rare. Most successful job information chains were one or two steps. Nicolas Christakis and James Fowler study influence and behavior change in human networks. They found that influence never travels beyond three steps and at three steps it is already a very weak signal (Christakis and Fowler, 2009). The strongest influence, in human networks, happens at one or two steps.
Everyone in a network is subject to a network horizon of approximately two steps. The horizon is unevenly distributed in our networks — sometimes it stops at 1 step and in rare cases in reaches out to 3 steps. Our view and influence into the network is limited. We might be a member of a large network, like a company with hundreds or thousands of employees, but that does mean we can access information, data, or advice when we need it. If our network horizon is limited, then we must be careful of how we build our network.
Another limitation of our ability to navigate our networks is time. Relationships in networks take time to build and then more time to maintain. This limits our networks — we can only have as many friends and colleagues as our time allows. Ronald Burt advises business people to avoid redundant ties (Burt, 1995) — connections to groups you already have connections to (via some other path). Burt shows that those who build their network wisely are more likely to be innovative (Burt, 2004) and more successful in their careers overall (Burt, 2000). It pays to build your network, and that of your group, wisely.
How groups and small communities (such as neighborhoods, NGOs, or departments in organizations) interact is also important to the success of the network. A fragmented network — with islands of information knowledge — is a death knell to both implementation and innovation in communities and organizations. Bridge building between natural islands of interest is a key step in laying the foundation for innovation. The exchange of diverse ideas, methods, and opinions allow new solutions to emerge.
Network efficiency can be looked at from two angles, both involve cost. One, is the cost building the network

— you usually want to have a minimum number of links to connect all of the available nodes. This results in a minimum spanning tree (Kruskal, 1956). Many man-made networks are trees — also known as hierarchies or hub-and-spoke networks. A tree network, that is cheap to build, is often expensive in delivery of data, information, or whatever travels across the network. Tree nets (hubs-and-spokes) are vulnerable — remove a hub or two and the network falls apart — no longer can many of the remaining nodes reach others. Business has always tried to get rid of redundancy in structures and processes. But redundancy actually helps networks. It provides multiple paths between most nodes and eliminates single points of failure. The right amount of redundancy establishes network resilience, without killing efficiency or cost.

Networks can be complex. As the number of nodes and their inter-connections grow — complexity increases in a rapid non-linear fashion. Pencil and paper can no longer help us once we get past a dozen nodes — which is most networks. To understand work networks, social networks, ecosystem networks, biological networks, economic networks, and community networks, we need the help of ICT.

Network Analysis Software

Network analysis, especially the layout algorithms and metrics, can get very complicated – a complexity that the average employee or community organizer cannot handle. Network analysis software can handle much of the complexity and the millions of calculations that networks of normal size require. The key to effective software is to have simple data inputs, hidden complexity of data processing, and easy to understand outputs. Any person that understands networks, and how to build them should be able to use simple software to support that effort. The software handles the mathematics and data base, while the consultant/analyst/builder handles the sociology.

Data for network analysis software is divided into two types: nodes and links. Node data tells us about the people/groups/organizations in the network — who they are, and their attributes. Node attribute data may include demographics, locations, group memberships or choices. For organizational employees we may track: gender, ethnicity, years of service, education, department, work location, or key skills. For a community member we may track: organizational memberships, skills, community roles, education/training, or experience. Although there is no limit to the number of attributes you can track for each node, we usually recommend to our clients to keep it to less than 10. Dividing people up into too many groups usually ends up confusing the analysis with too many possibilities. The rule of 7 +/- 2 is a good rule to follow for analysis. Humans can keep about 7 unique entities in their working memory at one time (Miller, 1956).

Even though a computer can easily handle more than 7 concurrent objects, our minds cannot. It is our minds, and their ability to see and understand patterns, that are key to making network maps useful. What the computer creates — a mathematically perfect layout, with many colors for many attributes — is useless if consultants/analysts/members/weavers of the network cannot understand it and make insights for action. We want the ICT to work as intelligent assistants for us, not present us information we cannot interpret and fit into our thinking.

The links in the network show us, who is connected to whom, and at what intensity. Employees can be connected to other employees in a variety of ways, from task ties, to advice/mentoring relationships, to purely social connections. Again, we follow the 7+/- 2 rule with links also. But occasionally there will be other interpersonal dynamics that determine the number of link sets — or networks — between the nodes. Corporate clients might get one network for each product line or one map for each major business process. NGOs might have multiple networks that cover regions, projects, clients, and services. One company had 26 product lines and 26 networks of how work got done to build and deliver each product. One Foundation produced 18 networks of their various grantees and the goals they worked on. Both organizations were eager to see who played a key role in an individual map and who appeared in multiple maps, connecting various processes, projects and goals.

Links show intensity and frequency of interaction. Corporate clients usually prefer to weight their network links using a Likert scale (Likert, 1932). A 5-point scale dividing interaction frequencies amongst "hourly, daily, weekly, monthly, more than once a year" is a common scale to examine business conversations. NGOs are usually not interested in such precise measures of interactions — they usually use a two category measure: regular/frequent and sporadic/light. Crime & corruption investigators and those probing other covert networks also use the latter scale, but a add a new attribute (actual/suspected). In covert networks not all links can be verified — criminals, corrupt politicians, and terrorists want to hide their connections (Krebs, 2002). Therefore observed/confirmed links are drawn with a solid link and potential/possible/suspected links are drawn with a dotted line.

Occasionally a few more attributes are added to link data. The most common added attribute is media-used during interaction. Sometimes it is as simple as dividing the communications between analog (Face-to-face, voice) and digital (computer-based). Other media attributes can reveal the timing of communication — synchronous (real-time: phone call) vs. asynchronous (distributed over time: email).

The easiest way to organize network data is a spreadsheet. One worksheet is for Node data — each Node in a separate row, and each attribute in a separate column. A second worksheet is for Link data — each tie/link in a separate row, with the first two columns showing From and To (which way the link goes). The later two

columns show frequency/strength of link and it's type or content. The data for a simple network is shown in Fig.4 below.

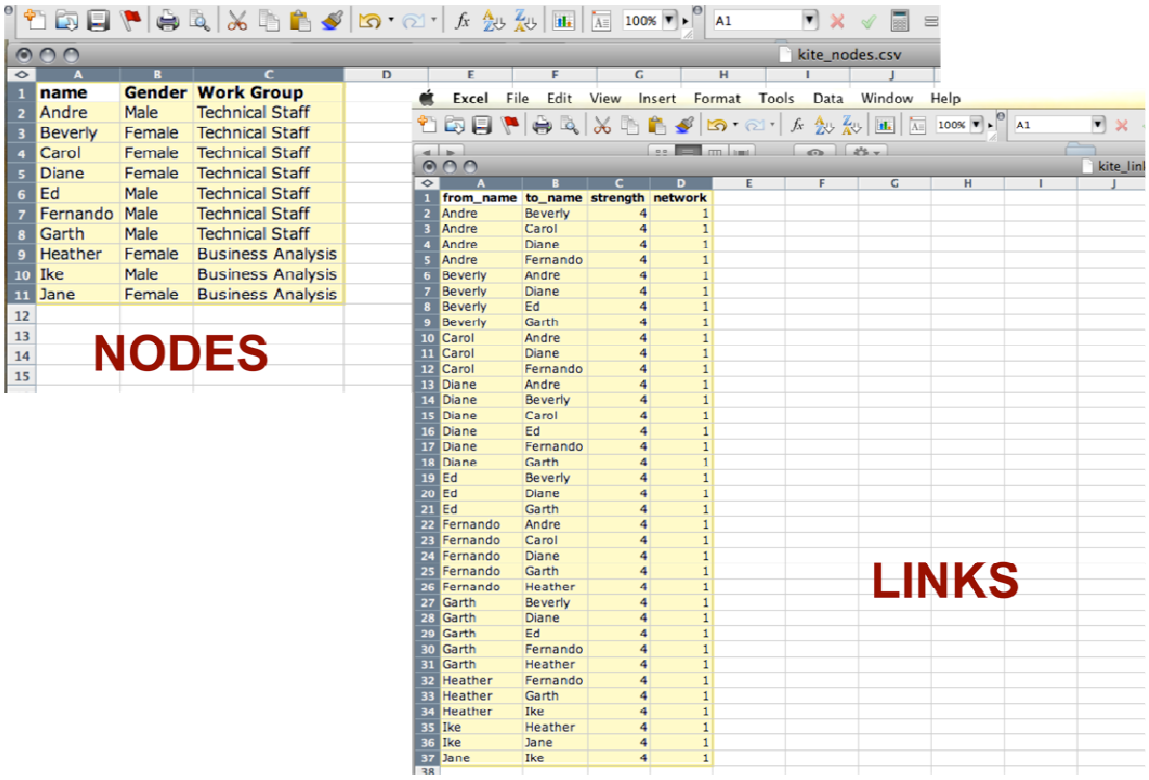


Fig. 4. **Spreadsheet of simple network data.**

Once the data is imported into the network analysis software, the user can arrange various layouts using the built-in network layout algorithms. It is advantageous if the software can keep track of multiple views of the network data. To properly evaluate a network the user/analyst needs to be able to quickly see various views/slices of the network — similar to how a doctor uses a CAT scan to study a patient. Along with a map of the sub-networks, we also need to have metrics of each view. Network metrics give us insights into various levels of the overall network. We can measure at the individual level — network centrality (Freeman, 1979). We can also measure at the group/cluster level — see how various attribute groups interact with each other and with the rest of the network (Krackhardt and Stern, 1988). Finally, we can measure the whole network — derive the weighted average path length, a good indicator of how efficient the network is moving information around.
Critical to this mapping and measuring process is the ability to "measure what is mapped" — the current network measure must reflect exactly what is drawn on the screen. The ability to change the network map — hide a node, add a link — and then measure the results of that change is important in doing "what-if" analysis. Corporations can model employee changes, or experiment with structural changes before committing to them. Community organizers can look at network weaving possibilities and how they will improve the network overall — in both resilience and efficiency. Investigators and crime-fighters can examine how best to break up a gang, criminal enterprise, or terrorist network — which node(s) should be removed to destabilize the network?
The current set of popular network metrics (Wasserman and Faust, 1994).
have a few drawbacks when measuring human networks. They make two dubious assumptions:

1) Information flows only along the shortest paths — geodesics;
2) Information flows along all geodesics — ignoring the network horizon.

The author has worked in more than 500 network analysis projects and has seen different dynamics in human networks. Information flows along more than the just the shortest paths in the network — even though you hear most information directly from your friends/colleagues, you often hear things about them from their friends/colleagues. We are not just connected one way to everyone else in our network — we are connected multiple ways, via multiple paths. The network horizon and influence research shows us that we should be focusing on the shorter paths in the network — those of lengths one, two and sometimes three. Therefore, network analysis software that focuses on mapping and measuring human networks should have the ability to measure information flowing along all paths that are within the network horizon. The Integration metric, which takes into account the network horizon and paths other than geodesics, is shown in Fig. 5 for each employee on the map.

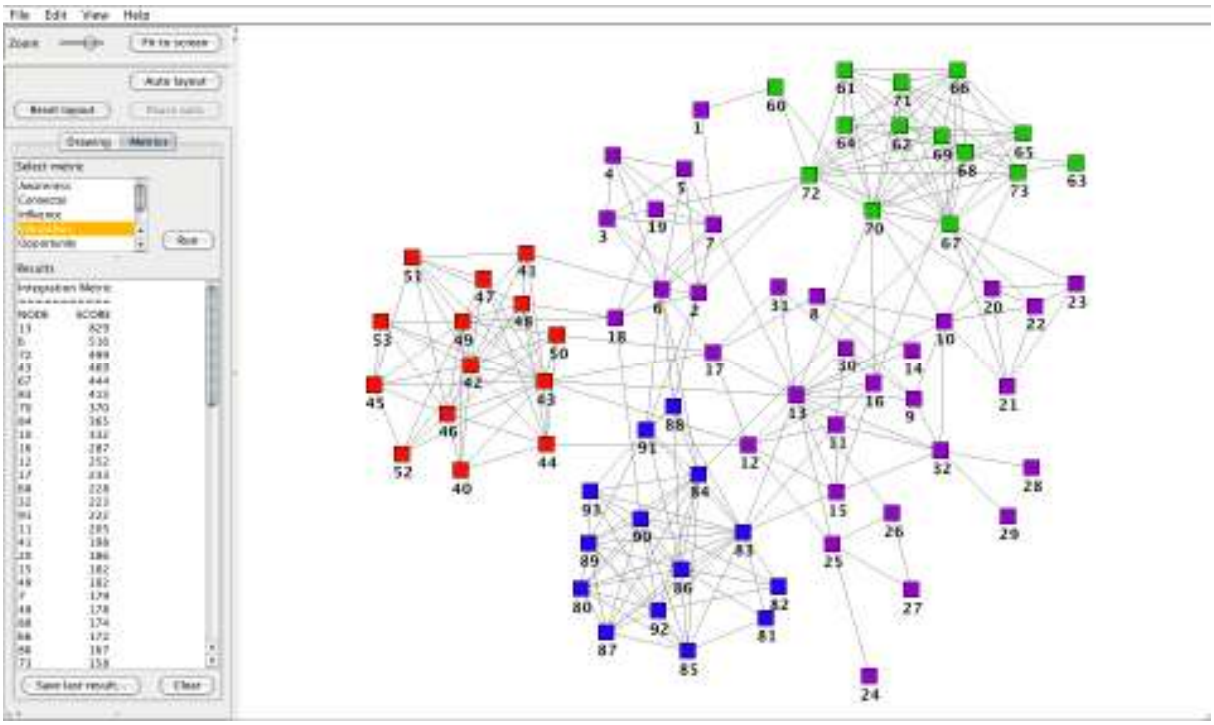


Fig. 5. **Software showing Network Map and Metrics.**

Above is a picture of a network analysis program, showing a small community of four organizations (each a different color) — each node is an employee of that organization. We see both Intra-organizational and Inter-organizational work ties — who works with whom on a specific regional project — employee names are replaced by numbers for privacy. The software shown exports the network map in SVG (Scalable Vector Graphics) format which can be used with desktop software or within web pages. SVG is a key component of the upcoming HTML 5 standard for web-based information display.

Useful network analysis software should also work well with other popular software such as spreadsheets, databases, word processing programs, presentation programs, and web design tools. Popular and easy-to-use data formats should be utilized in the exchange of data between programs. Comma-separated Values (CSV) is still the easiest way to exchange data between computer programs. Extended Markup Language (XML) was popular for a while with web-based data, but is now losing popularity because of the need to write data-extraction programs to move data between one source and another. CSV files are fully supported in popular business software that is found on Linux, Macintosh and Windows computers.

Graphic file formats are critical to network mapping software. Many programs only export bit-mapped formats such as *.bmp, *.gif, and *.jpg. These formats tend to lose their quality and readability when re-sized in reports or presentations. Network analysis programs should support popular vector graphics formats such as *.pdf, *.svg, and *.emf. These network maps can be re-sized, and often re-edited outside of the original network analysis software. Network maps are often edited for delivery and presentation in vector graphics programs such as Adobe Illustrator and Open Office - Draw.

Conclusion

Network analysis software supports the end-user in knowing the network, while concurrently supporting the network weaver in knitting the network. With an accurate social map of how things happen, those who want to navigate the network, or nudge/influence the network now have an "as is" diagram and lists of key leverage points (via the metrics).

Over time the network changes structure and shape. The network software allows for showing minor and major changes, along with reporting updated metrics. A good network analysis software package should be similar to Google Maps on the WWW or your smartphone — you always know where you are currently, you can go back to where you came from, and you can zoom in on your destination. Network maps will be saved and printed, or sent to a friend via email. The maps will be edited further using various graphics programs and displayed on the desktop or WWW. Whether we are looking at geographic maps or social maps, they guide us from where we are, to where we want to go — and document the trip along the way.

Acknowledgements

Supported by ERAF project “Visualization and analysis tools’ set of dynamic networks”
(No. 2010/0318/2DP/2.1.1.1.0/10/APIA/VIAA/104) at The Institute of Mathematics and Computer Science,

University of Latvia.

References

Burt, R., 1995. *Structural holes: The social structure of competition*. Harvard University Press, 1995.

Burt, R., 2004. Structural holes and good idejas. *American journal of sociology* 110.2 (2004): 349-399.

Burt, R., 2000. The network structure of social capital. *Research in organizational behavior* 22 (2000): 345-423.

Christakis, Nicholas, A., and Fowler, James, H.,2009. *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown, 2009.

Freeman, Linton, C.,1979. Centrality in social networks conceptual clarification. *Social networks* 1.3 (1979): 215-239.

Friedkin, N., 1983. Horizons of observability and limits of informal control in organizations. *Social Forces* 62.1 (1983): 54-77.

Granovetter, M.,1995. *Getting a job: A study of contacts and careers*. University of Chicago Press, 1995.

Krackhardt, D., and Hanson, J.R.,1993. Informal networks. *Harvard Business Review* 71.4 (1993): 104-11.

Krackhardt, D. and Stern, R.N.,1988. Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly* (1988): 123-140.

Krebs, V.,2002. Mapping networks of terrorist cells. *Connections* 24.3 (2002): 43-52.

Krebs, V.; 2005. Holley, June. "Building adaptive communities through network weaving". *Nonprofit Quarterly* vol. 12 (Winter 2005): 66-72.

Krebs, V.,2007. Managing the 21st The Century Organization. *IHRIM Journal* 11.2 (2007): 2-8.

Kruskal, J., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society* 7.1 (1956): 48-50.

Likert, R.,1932. A technique for the measurement of attitudes. *Archives of psychology* (1932).

Miller, G.,1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 63.2 (1956): 81.

Wasserman, S. and Faust, K., 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge University Press, 1994.

Watts, Duncan, J., 2004. *Six degrees: The science of a connected age*. WW Norton, 2004.

Different solutions of MySQL in the cloud – security and possibilities

Zigmunds Bulins, Vjaceslavs Sitikovs

Institute of Applied Software, Riga Technical University, Meza 1/3, Riga, LV-1048, Latvia
Zigmund.Bulinsh@gmail.com, Vjaceslavs.Sitikovs@rtu.lv

Abstract: *Cloud computing is a good way to raise productivity of offered service without investments into new infrastructure, training of the personnel or software acquisition. This technology expands potential possibilities of existing information systems. In recent years cloud computing grew from good business concept to one of the most demanded industry in information technologies. The paper contains a short review of different cloud database providers which uses MySQL as a basis. Technical nuances, potential problems and risks related to migration of the existing MySQL databases to the new environment are reviewed. In the paper we try to review the actual possibilities of the new platform and compare the cloud DBaaS (database as a service) solutions which are implemented with MySQL database management system widely used in the web.*

Keywords: MySQL, DBaaS, Heroku, Google Cloud SQL, ClearDB.

Introduction

Cloud computing is a dynamic method of increasing productivity of service or possibilities without investments in new infrastructure, training of the personnel or software licensing. This expands possibilities of existing information systems. In recent years cloud computing grew from good concept business to one of the most quickly developing industries of information technologies (Chandra, Mondal, 2011).
In the last two years there was quite a lot of activity around the cloud databases on the stage – Google, Amazon and Xeround (Xeround, 2012a) companies announced their DBaaS based on the MySQL database. We will review three of them – Google Cloud SQL (Google, 2012b), Xeround, ClearDB (ClearDB, 2012).

The Google Cloud SQL

Google Cloud SQL is a web service which allows creating, forming and using relational databases with App Engine (Google, 2012a) applications. This is completely self-managed service which supports and manages databases, allowing developers to concentrate on implementation of applications and necessary services.
Offering functionality of the MySQL database, service allows moving easily the data, applications and services to the cloud and out of it. It allows increasing mobility of data and provides faster entering the market because there is an ability to quickly scale an existing database.
To guarantee service availability for critical applications and services, Google Cloud SQL replicates data in different geographical areas for ensuring high availability of data.
Main features of the Google Cloud SQL service are (Google, 2012b):

- Ease of use – a rich graphical user interface allows for creating, configuring, managing, and monitoring the database instances;
- Fully managed – no worrying about tasks such as replication, patch management, or other database management chores - all these tasks are provided by “cloud”;
- Highly available – to meet the critical availability needs of today's applications and services, features like replication across multiple geographic regions are built in, so the service is available even if a datacenter becomes unavailable;
- Integrated with Google App Engine and other Google services – make it possible to work across multiple products easily, get more value from the data, move the data into and out of the “cloud”, and get better performance.

If we compare the Google Cloud SQL with the others similar services on the market like Amazon EC2 (Amazon, 2013), Windows Azure (Microsoft, 2013), Xeround then two possible models of the DBaaS service are met:

- Virtual images of the configured database instances which are running on the virtualized hardware;
- Distributed and automatically managed database which is not linked to specific location in the cloud.

With Google Cloud SQL we get the second option.

Limitation of the Google Cloud SQL

The Google Cloud SQL represents MySQL DBMS placed in a “cloud”. Google Cloud SQL provides all functions which are offered by MySQL DBMS, but with several limitations. There are the following main restrictions of cloud service (Google, 2012b):

- The size of a separate instance of a database is limited by 10 gigabytes;
- User defined functions (UDF) are not supported;

- Replication functionality is not available for configuring and setting up;
- File based functions are blocked (such as DATA INFILE, LOAD_FILE etc.).

As the size of an instance is limited this service is applicable generally only for applications of small and medium business (as it was described in the Google Cloud SQL service description). Also this restriction can concern multimedia and similar applications which store binary data in a database and, as a rule, it uses a lot of disk space.

Absence of replication support is the minor shortcoming as Google Cloud SQL itself implements this mechanism and tracks its correct operation.

One more restriction which isn't specified is an absence of federated table support. If migrated application that already is using this mechanism, it is necessary to alter architecture of application in case of transfer it to Google Cloud SQL platform. Most likely the new platform automatically solves the problem because of which remote tables were used. If not, it will be necessary to find other solution within the new environment.

One more technical restriction of cloud DBMS is that the platform is intended for applications with a low to average level of intensity of data recording (Google, 2012b). That is, applications with intensive data recording will not work effectively in the environment of Google Cloud SQL as the replication engine is used.

In DBMS instance access control there is a new layer – a Google API project layer. Knowledge of DBMS access codes is not enough for acquiring access to a database. It is necessary to have also access to the project within which DBMS was created. The project and access to it are managed by the service of Google API Console (Google, 2012d) in a Google account.

Available tools for the service interaction

Unlike stand-alone MySQL DBMS, in Google Cloud SQL there is no possibility to connect to the database directly from any computer. Connection is carried out or by means of the web browser, or using the special Command Line Tool program.

In addition it is possible to use the SQuirrel SQL application (Universal, 2012), which actually uses the aforementioned program tool to execute commands on Google Cloud SQL (Fig. 2).

In order to connect to the Google Cloud SQL database from a certain computer, it is necessary to generate an access key in settings of the service account and later submit it to the program tool on first use – this mechanism allows the recipient of service to supervise access to instances of databases.

In a given context threats relates to the service provider and not to the Google Cloud SQL user, but that's doesn't free user from potential risks.

The question of trust consists in issue that data takes place in the environment belonging and supervised by the third parties where there is no possibility to track their actions. That is not applicable for public government institutions or systems with confidential data (changes in the way how data is stored in the DBMS are necessary, to use Google Cloud SQL environment).

Connection is carried out or by means of a web browser (Fig. 1), or using the special Command Line tool (Fig.2).

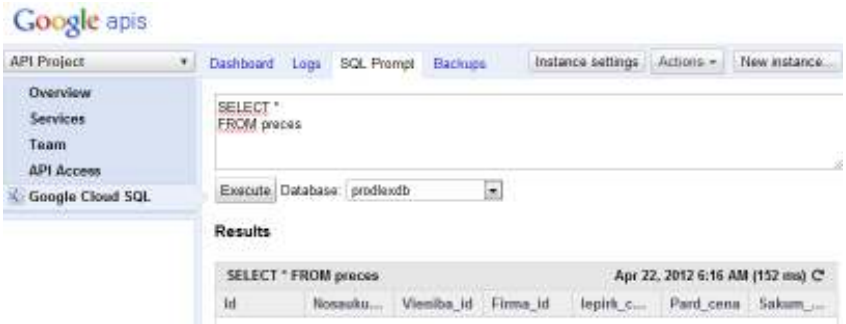


Fig. 1. **Web browser access**.

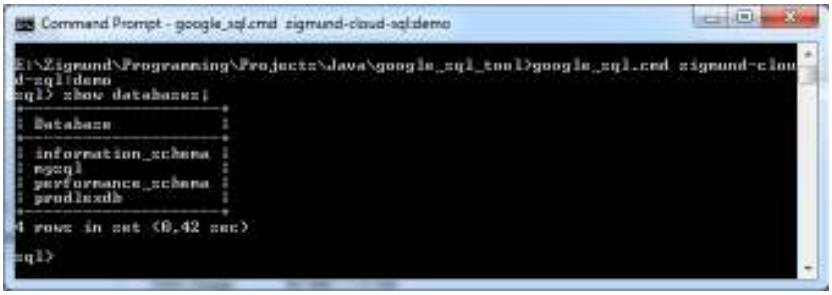


Fig. 2. **Command line tool**.

Xeround

In contrast to Google Cloud SQL, which actually uses the almost not changed MySQL version 5.5 (no modifications were made, just some features were disabled), Xeround is built on the MySQL Storage Engine Architecture, acting as a pluggable storage engine (Fig. 3). Relying on this architecture, and the MySQL query language support, Xeround patented storage engine seamlessly replaces current MySQL database (Xeround, 2012b).

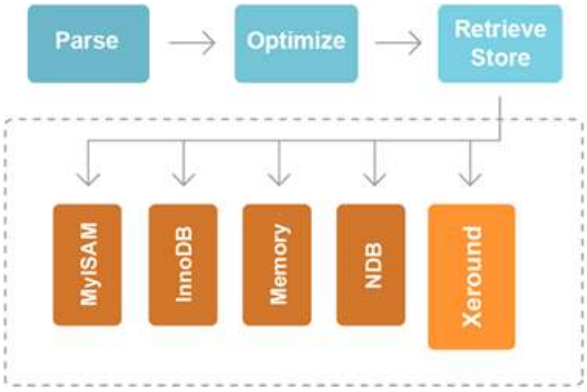


Fig. 3. **Xeround structure** (Xeround, 2012b).

Xeround's two tier architecture is comprised of Access Nodes and Data Nodes. Data Nodes are responsible for storing the data, while Access Nodes receive application requests, communicate with Data Nodes, perform computations and deliver request results. Xeround stores data in virtual partitions that are not bound to the underlying hardware infrastructure. Each partition is replicated to the different Data Nodes located on separate servers, providing high availability and full resiliency (Xeround, 2012b).

The background of the Xeround cloud database service is the MySQL server of version 5.1 (as of January 2013, using our test account). This actually is not widely advertised on the company's website, but for some certain segment of customers that can be an important issue as there are quite many improvements in the SQL syntax, optimizer and other places of the server (MySQL, 2012).

But the main advantage over the Google Cloud SQL is that Xeround database is not vendor locked – can be run on any cloud platform and any stack (Xeround, 2012b).

It is possible to connect directly to the Xeround cloud database using any tool available on the market – there is nothing to describe in contrast to Google Cloud SQL which uses its own tools.

ClearDB

Another available option for the cloud database as a service is the ClearDB which can be used on different cloud platforms such as Heroku (Heroku, 2012) or Amazon (Amazon, 2013). ClearDB is similar to the Xeround solution, however it does not invent new storage engine, but instead enables scaling and durability for the unchanged MySQL functionality.

ClearDB creates "multi-master" and "multi-master with multi-replica" MySQL configurations in geographical regions that are important to the customer to provide applications with a fully redundant solution that can survive outages, network failures and even natural disasters (ClearDB, 2012).

ClearDB uses a combination of advanced replication techniques, advanced cluster technology, and layered web services to provide you with a MySQL database that is "smarter" than usual. We also use things like mixed binary replication logging and auto-increment offset seeding so that it is possible to continue using MySQL's non-determanistic and time-based functions such as UUID(), NOW() as well as auto-increment keys in the tables (ClearDB, 2012).

With the ClearDB solution it is also possible to connect to the database using any available tool as in case of Xeround.

There are also few feedbacks found on the internet about the issues with Xeround DBaaS, when users switched to ClearDB in place of the first one because of the query execution issues (Stack Overflow, 2012) – this can be a positive sign for the ClearDB.

Technological risks in the cloud

With growth of number of users and companies which store the data in "clouds", more frequently the questions about safety come up (Subashini, Kavitha, 2010). Despite all activity around cloud computing, business sector's clients still don't wish to place their systems to the cloud environments. Security is the main reason which detains rapid development of the market of cloud computing (Marstona et al 2011). Also questions of a privacy of data and problem of its protection continue to influence the market of cloud computing (Subashini, Kavitha, 2010; Mansfield-Devine, 2008). Recent IDCI survey shows that 74% of technical directors and managers of

information technology sector noted security as the main challenge which keeps them from adoption of existing systems to model of cloud service (Subashini, Kavitha, 2010).
In the case of database migration to the environment of a Google Cloud SQL there are new security (Fig. 4) and trust challenges for users. Security challenges are generally related to new elements of infrastructure and the platform.

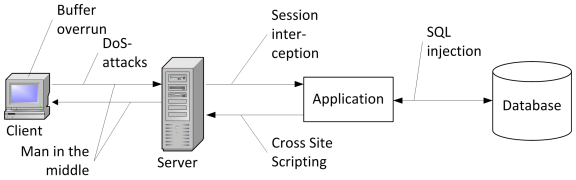


Fig. 4. **Threats by location**.

Cloud computing threats can be grouped in 5 main classes (Fig. 5) as described farther.
Functional Threats of Cloud Components
This type of attack is associated with multiple layers of the "clouds", the main principle of security is that the total level of security is determined by the security of the weakest element (Subashini, Kavitha, 2010). So, denial-of-service attack (DoS attack) on a proxy-server setting in front of the cloud will block access to the whole "cloud", despite the fact that all works smoothly within the "cloud".
In a similar way, SQL injection, which occurs on the application server, will provide access to data storage systems, regardless of access rules in data storage layer (Korzhov, 2010).

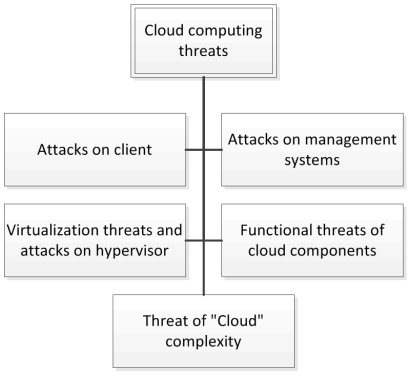


Fig. 5. **Simple threats classification**.

Attacks on a Client

These types of attacks have worked out in a web environment, but they are just as relevant in cloud environments as users connect to the cloud through a web browser. Attacks include such types as Cross Site Scripting (XSS), DoS attacks, interception of web sessions, stealing passwords, "the man in the middle" and others. Protection against these types of attack is traditionally a strong authentication using an encrypted connection with mutual authentication. But not all creators of "clouds" can afford such expensive and often inconvenient means of protection (Chonka et al 2011).

Virtualization Threats and Attacks on a Hypervisor

In IT, the hypervisor, in a different way is called "virtual machine manager".
Since the platform for the cloud elements, usually is a virtual environment, the attack on virtualization threatens the entire cloud as a whole (Lombardia, Di Pietro, 2011). This type of attack is unique to cloud computing (Rosenthal et al, 2010).
At the moment, there are known few real attacks on hypervisors, but it is possible that the amount of such attacks will rise in the future (Rutkowska, Tereshkin, 2008).

Threat of a "Cloud" Complexity

Monitoring the events in the "cloud" and management of them is also a security issue. How do we ensure that all resources are counted and that there is no rogue virtual machine that perform third-party processes and do not interfere in mutual configuration of the layers and elements of the "cloud" (Kritsonis, 2011)?
This type of threat associated with the processing of the cloud as a whole and the search for fraud and other irregularities in the "cloud" structure, which can lead to unnecessary expenditure on maintaining the "health" of the information system (Korzhov, 2010).
The level of this type of threat is the highest, and it is assumed that it is impossible to create a universal remedy to protect against that - for each "cloud" individual protection system must be built.

Attacks on a management systems

A large number of virtual machines that are used in the "clouds", especially in public clouds, require a management system that can reliably control the creation, transfer and utilization of virtual machines. The interference in the management system can lead to ghost virtual machines, blocking some of the machines and the substitution of elements or layers in the cloud to the rogue. All this allows an attacker to gain access to the data of the "cloud" or to gain control over part or the whole "cloud" (Paquettea et al, 2010).

The comparison

To summarize the short reviews of the DBaaS solutions we will present the comparison table, which contains several criteria (Table 1).

Table 1

Simple comparison of the DBaaS.

Solution	Google Cloud SQL	Xeround	ClearDB
Base MySQL version	5.5	5.1	5.5
Uses replication	Yes	No	No
Implemented as a storage engine	No	Yes	No
SSL support	Yes (only platform internal connections)	Yes	Yes
Direct external connections	No	Yes	Yes
Is tied to a certain platform	Yes, Google	No	No
Can be easily moved to another provider/platform	No	Yes	No

Conclusion and future work

With release of the new Google Cloud SQL service providing the cloud version of MySQL 5.5 DBMS and other DBaaS such as Xeround and ClearDB, many new possibilities appear for migration of existing systems to the new architecture with rather small losses upon transition. Carrying out the analysis of possibility of transition to cloud model of a database, it is necessary to consider some risk factors described in this article and to ensure that existing system corresponds to restrictions of service and won't exceed them soon.

Transition to a cloud database gives high scalability of system in context of accessibility as the supplier of service provides distribution of data between many geographical locations that accelerates data access to the end users.

There are different cloud databases which are implemented on the basis of MySQL server distribution, but each of them uses different technique to implement the cloud platform support. If we talk about Google then we are tied to a certain platform, a single provider and also some technical limitations (which were described in this article).

If you are searching for a simple solution, which means just to move the data to the cloud database without any changes – then ClearDB is right choice, as it provides the original and latest MySQL 5.5 version in the cloud.

Xeround solution is different from the ClearDB technically (as described in sections before), but it has a possibility to move data from one cloud provider to another one – what can be important if geographical location of the data is important for the company or application (it's host agnostic, you can easily move your data anytime).

Acknowledgements

This work was funded by the ERDF (ERAF) project No. 2011/008/2DP/2.1.1.1.0/10/APIA/VIAA/018 "Development of Insurance Distributed Software Based on Intelligent Agents, Modelling, and Web Technologies" and we would like to acknowledge that support.

References

Amazon, 2013. Amazon Elastic Compute Cloud (Amazon EC2). Available: http://aws.amazon.com/ec2/. Last accessed 20th Jan 2013.

Chandra, S.M., Mondal, A., 2011. Identification of a company's suitability for the adoption of cloud computing and modeling its corresponding Return on Investment. In Mathematical and Computer Modeling. 53 (3-4), 504–521.

Chonka, A., Xiang, Y., Zhou, W., Bonti, A., 2011. Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks. In Journal of Network and Computer Applications. 34 (4), 1097–1107.

ClearDB, 2012. ClearDB –The Safer, Reliable MySQL Cloud Database For Your Applications. Available: http://www.cleardb.com/better.view. Last accessed 31 Jan 2013.

Google, 2012a. Google App Engine. Available: https://developers.google.com/appengine/. Last accessed 20th Jan 2013.

Google, 2012b. Google Cloud SQL. Available: https://developers.google.com/cloud-sql/. Last accessed 20th Jan 2013.

Google, 2012c. Command Line Tool. Available: https://developers.google.com/cloud-sql/docs/commandline. Last accessed 20th Jan 2013.

Google, 2012d. Dashboard. Available: https://code.google.com/apis/console. Last accessed 20th Jan 2012.

Heroku, 2012. Heroku - Cloud Application Platform. Available: http://www.heroku.com/. Last accessed 31 Jan 2013.

Korzhov, V., 2010. Clouds: Legends and Myths (in Russian). Available: http://www.anti-malware.ru/node/2333. Last accessed 20th Jan 2013.

Kritsonis, T., 2011. Security Risks in the Cloud – Reality, or A Broken Record? In Infosecurity. 8 (1), 20-23.

Lombardia, F., Di Pietro, R., 2011. Secure virtualization for cloud computing. In Journal of Network and Computer Applications. 34 (4), 1113–1122.

Mansfield-Devine, S., 2008. Cloud Security: Danger in the clouds. In Network Security. 2008 (12), 9-11.

Marstona, S., Lia, Z., Bandyopadhyaya, S., Zhanga, J., Ghalsasib, A., 2011. Cloud computing — The business perspective. In Decision Support Systems. 51 (1), 176-189.

Microsoft, 2013. A rock-solid cloud platform for blue-sky thinking. Available: http://www.windowsazure.com/en-us/. Last accessed 20th Jan 2013.

MySQL, 2012. MySQL 5.5 Reference Manual: What Is New in MySQL 5.5. Available: http://dev.mysql.com/doc/refman/5.5/en/mysql-nutshell.html. Last accessed 20th Jan 2013.

Paquettea, S., Jaegerb, P.T., Wilson, S.C., 2010. Identifying the security risks associated with governmental use of cloud computing. In Government Information Quarterly. 27 (3), 245–253.

Rosenthal, A., Mork, P., Hao Li, M., Stanford, J., Koester, D., Reynolds, P., 2010. Cloud computing: A new business paradigm for biomedical information sharing. In Journal of Biomedical Informatics. 43 (1), 342–353.

Rutkowska, J., Tereshkin, A., 2008. Bluepilling the Xen Hypervisor. Available: http://invisiblethingslab.com/resources/bh08/part3.pdf. Last accessed 20th Jan 2013.

Stack Overflow, 2012. heroku – Xeround vs ClearDB hosted MySQL pros and cons?. Available: http://stackoverflow.com/questions/10099239/xeround-vs-cleardb-hosted-mysql-pros-and-cons. Last accessed 31 Jan 2013.

Subashini, S., Kavitha., V., 2011. A survey on security issues in service delivery models of cloud computing. In Journal of Network and Computer Applications. 34 (1), 1-11.

Universal, 2012. Squirrel SQL Client Home. Available: http://squirrel-sql.sourceforge.net/. Last accessed 20th Jan 2013.

Xeround, 2012a. DBaaS – 'Worry-Free' Cloud Database for MySQL Applications. Available: http://xeround.com/mysql-cloud-db-overview/dbaas/. Last accessed 20th Jan 2013.

Xeround, 2012b. MySQL scalability and high availability - cloud DB from Xeround. Available: http://xeround.com/mysql-cloud-db-overview/. Last accessed 20th Jan 2013.

E-health progress in Latvia

Linda Butane, Nauris Paulins

Latvia University of Agriculture, Faculty of Information Technologies, Liela Str.2, Jelgava, LV-3001, Latvia
butane.linda@gmail.com

Abstract: *The implementation of e-health system in Latvia healthcare has been initiated since 2003, but for long time there was no success in this area. Therefore this paper analyzes e-health project progress in Latvia, implementation process and current situation. To understand main problems of so problematic e-health implementation, in this paper analyzed national strategic documents and normative documents as well as several informative resources. The last document, approved in 2010 includes 3 project development parts witch should be implemented to year 2013: 1) Electronic health card and integration platform solution information system; 2) Electronic appointment booking system, healthcare workflow digitizing, society health portal; 3) Electronic prescriptions information system. Currently Latvia is at start of e-health implementation second part process for e-health program and systems development. This paper analyzes current status for e-health information system development process as well as discovers main e-health implementation problematic: interoperability, maintenance, legal and regulatory restrictions and security and confidentiality of patient's data. Currently the main topics of the discussion are associated with data security and confidentiality.*

Keywords: e-health, e-health progress, e-health implementation, e-health in Latvia.

Introduction

The main objectives of Latvia e-health development are to improve health care system, promote individuals of their health and reduce time waste in information collection and access. In European Union e-health is identified as a key tool for a health care quality, access and safety promotion. Latvia is one of the EU27 member states where e-health is used only to a limited extent. Almost last 10 years there were several attempts to reform healthcare system in Latvia and implement e-health in Latvia healthcare, but for long time there was no success in this area. In last year's communication about e-health is not so active anymore, like it was couple years ago. Currently is hard to understand main problems of so problematic e-health implementation, because IT usage and broadband connection for Latvia is very good. Communication infrastructure is highly developed, according Latvian Information and Communication Technology Association data, Latvia internet has 4th highest upload speed and 7th highest download speed. The same time the Ministry of Health has developed national e-health policy/concept document, which has been approved by the Cabinet of Ministers of Latvia on August, 2005. Guidelines provide for the development of e-health in Latvia through investments totaling 30 million EUR over the period of 10 years. (Technology Development Forum, 2011) But it seems that there is more problems than solutions, for example in 2011 in public area were active discussion about to stop current process and overtake e-health from Estonia. There already was investment from European Regional Development Fund (ERDF) (2007-2013) budget for e-health system development in Latvia although none of information system is launched yet (Springe, 2011). In the same period of time Estonians managed to develop and implement e-health solution that provides the main necessary services to residents. Aim of this research is to analyze current e-health progress in Latvia and discover main problems in this area and discover future trends of analysis for e-health implementation in Latvia.

Materials and methods

Materials and methods include analysis of normative documents and several informative sources. Data presented are mostly based on national strategic documents, governmental reports and other related institution publications and conferences presentations.

Results and discussion

E-health can be defined as the use of information and communication technologies (ICT) in the field of healthcare. E-health is existing and also new e-services development and improvement in the sector of healthcare. E-health solutions significantly would improve the health care industry and the methods of traditional approaches - mainly in the field of information processing. In Latvia e-health information system (IS) is developed in order to improve health care efficiency and quality.
The major benefit of e-health is better health care quality, doctors will have access to detailed information about the patient: case history, made manipulations, used medicine. Another benefit: efficiency and rational use of resources by avoiding duplicate tests already taken, diagnosis made. This means that the patients will faster receive diagnosis and treatment.

The main objectives of e-health development determined by National Health Service (E-health Latvia, 2012) are to improve health, promote individual control of their health; reduce wasted time spend on patients contacts with medical institutions; increase the effectiveness of the health care, providing health care specialists with a quick access to necessary patient health data; reduce the amount of information that health care specialists need to enter into the documents; increase the amount and usability of a structured information; increase effectiveness of medical institutions; increase health care data reliability and security.

E-health projects are supported by the Council of Europe, as the e-health has been identified as one of the major European Union priorities in the health care industry. E-health is identified as a powerful strategy to transform the health care system for a health care quality, access and safety promotion, enabling individuals and communities to enhance prevention, diagnosis, treatment, monitoring and management etc. of health.

E-health systems are implemented in many countries and all opinions on its effectiveness are more positive than negative. E-health will improve the quality of health care, control, statistics, will provide doctors with fast access to the required information (e.g., emergency call or difficulty in contact with a patient case), and will make it possible to control the queue length. E-health will improve control of the movement of financial resources, transparency - ability to track data flow, etc. (Cipule, 2012)

The following socioeconomic researches (Gartner Group, 2009) benefits have been identified as a direct result of adoption and usage of the e-health:

- 22% gain in clinical staff productivity,
- 15% reduction in prescription error,
- 16% reduction in waiting times for first outpatient appointment,
- 48% reduction in duplicate laboratory/chemistry tests,
- 99% reduction in loss of radiology images, etc.

E-health project implementation is connected also with other sectors – welfare, structures of the interior and justice system, as a certain amount of data could be obtained, for example, social services for work with children from disadvantaged families, as well as the court, if any of the process participants will not be able to ensure their presence in court of health problems. (Cipule, 2012)

E-health information system implementation process

Almost last 10 years there were several attempts to reform healthcare system in Latvia and implement e-health in Latvia healthcare, but for long time there was no success in this area. In Latvia the Ministry of Health is the main organization responsible for the development of national e-health policy and related implementation plans. The guidelines "*e-Veselība Latvijā*", was approved by the Cabinet of Ministers of Latvia on August, 2005. This document involves strategic advice for the development of information and communication technologies in the sector of healthcare up until 2013. The action plan for implementation of the priorities for year 2008-2010 was accepted in October, 2007.

The main tasks defined in this action plan (WSIS, 2013) are:

- centralized data storage and access to the patient's medical data;
- possibility to prepare an integration platform providing for safe information exchange between the health sector and information systems of other sectors;
- development of electronic prescription information system and safe data transmission channels between main information systems within the health sector;
- unified emergency medical assistance and disaster medical services information system.

Responsible organization for e-health implementation in Latvia is the National Health Service (NHS).

Latvian e-health information system implementation has begun similar as in other European Union countries as e-health project implementation of delivering e-health services for citizens and medical professionals.

- Since September, 2010 (Technology Development Forum, 2011) the first pilot projects e-health services have been developed within the first e-health pilot project and made available online at the Latvian unified government e-service portal Latvija.lv (https://www.latvija.lv/). The e-health services include:
- "My state-funded health care services" - the users can find information on their own or their children's medical visits, diagnoses and use the current Health Payment Centre's database.
- "My data in the register of diabetic patients" - this service provides information that is stored electronically in the register of patients with diabetes mellitus.
- "My newborn baby data" - this service provides information that is electronically stored in the register of neonatal information system.
- "My family doctor" - the users can get information on their own or their children's family doctor.

At the end of 2009 have been initiated three projects co-funded by European Regional Development Fund. In late 2010 went procurement procedures for all three projects within the implementation information system development. E-health information system development project was initiated in early 2011.

In Latvia e-health projects development process has started later as in other counties, also neighbor country Estonia where e-health implementation process was initiated in early 2000 and implemented since year 2005.

Exactly so long Estonians are already using their e-health system. There is no one certain reason why e-health development process in Latvia was initiated late. In society is listed the health sector untenable willingness to accept it, since there are now many other topical issues - salaries, medicines, also finances, etc. Against this background, e-health remains in second, third plan. (Ričika, Zonne, 2006) This is not a medical issue, but the political will, legal regulation and financial support and health care staff attitude for the issue. Medical institutions and general practices (GP) should be able to create electronic registration systems. Considering the complexity of the sector, it is important to have unified settings. Estonians have electronic identity (ID) cards already introduced and operating successfully - it is closely related to e-health. In Latvia, unfortunately, everything comes more slowly - identification cards, secure electronic signature, and it is affecting other electronic system development and implementation. (Zvejniece, 2011)

In Estonia, a number of e-health implementation positions are funded from the state budget, Latvia was expecting for ERDF funds. The health sector restructuring also delayed project implementation because of institutions change involved in the project.

According to guidelines „*e-Veselība Latvijā*" that determines the strategic directions of ICT development in health care, e-health information system project in Latvia is planned to implement to year 2016.

E-health information system development process

E-health implementation process is separated in three main parts in Latvia. In the first part will be developed a public health portal, in which will be available three e-services: e-booking, electronic health records and e-prescription. In the projects first part will be developed also an environment where the exchange of information between the various health information systems will be provided, as well as information security and data protection solutions developed. According to e-health project development hierarchy plan (Gabaliņa, 2012), prepared by the Ministry of Health, the following steps are scheduled:

- pilot e-services;
- e-booking and e-referrals;
- e-health portal;
- electronic health records (EHR);
- e-prescription;
- the industry supervision (unified monitoring information system (VUIS));
- industry standardization initiation;
- interoperability solution (integration platform (IP));
- ensuring the interoperability of existing IS, process changes;
- health care services accounting and payment system (DRG) model integration for stationary funding.

Currently are being developed three e-health projects within the first part of e-health information system development project "*E-veselības integrētās informācijas sistēmas attīstība*" (VEC e-veselības projektu arhitektūras vadlīnijas, 2010), which should be implemented to the year 2013:

1. Electronic health card and integration platform solution (IP) information system development and implementation 1^{ht} part

Integration platform development aim is to create environment which will ensure the exchange of information between the various health information systems. Integration platform is to present the e-health system users (people and systems) uniform and uncontroversial vision of centralized health care data, as well as provide shared technical functions of all central e-health information systems. Integration platform to provide the following function groups are set:

- classification of distribution (Classifier module);
- messaging treatment (Message Processing Module);
- personal data processing audit events (Data Audit module);
- e-health authorization (E-Health Authorization module);
- integration platform alert (notifications) processing (warnings module), etc.

Electronic health record system aim (VEC e-veselības projektu arhitektūras vadlīnijas, 2010) is to create a single unified database, where will be stored the data about provided services for patients in one place to ensure that these data are available for both the patient and medical staff from different medical institutions. EHR system provided functions conceptually are separable into two groups: EHR records management functionality, which includes medical records storage, access, and accuracy control of the data, as well as the use of EHR modules practical functionality using the EHR and data available in other e-health systems with strictly set EHR extensions (modules) interface.

2. Electronic appointment booking system development (e-booking), healthcare workflow digitizing (e-referral) – 1th part, society health portal development (e-health portal), information security and ensuring the protection of person's data.

E-booking information system provides (VEC e-veselības projektu arhitektūras vadlīnijas, 2010) workflow automation patients sequence listing with a doctor, based on the patient's initiative and/or expert's patient

referrals. E-booking financial module provides a general and universal mechanism for processing requests for payment of all medical institutions.

The project objectives are (Nacionālais veselības dienests, 2012):

- to provide an unified approach to information security and the protection of personal data in the health sector;
- to ensure effective financial management in the health sector;
- to increase patient awareness of health issues and the motivation to keep up their health. Create a health portal to provide patients and physicians access to personalized information and electronic services;
- to increase the availability of health care services;
- to increase physician productivity.

E-health portal will be available to doctors, pharmacists for work and also to patients and citizens for possibility to use e-services at the national e-government portal where recording system will be implemented, opportunity to clarify the state-funded services, facilitate data submissions, electronically advices. In health records will be imported information from the Centre of Health Economics and the Health Payment Center databases. Already, within a pilot project, part of the information stored in the databases can be viewed online in portal "https://www.latvija.lv/". Online authorization by a secure electronic signature, mobile ID authentication or accessing through the internet bank anyone can view profile data about themselves and their minor children using proposed 4 e-services. (Zvejniece, 2011)

3. Electronic prescriptions (e-prescriptions) information system development 1th part.

E-prescription information system provides conventional and specialty electronic prescription movement.

The project objectives are (Nacionālais veselības dienests, 2012):

- to ensure effective prescribing process for doctors;
- to provide more information to doctors about how patients are taking medications;
- to provide physicians and patients with a complete and objective information on patient medication;
- to ensure better monitoring and information on drug movement for industry management;
- to provide effective financial management in pharmacies.

The planned necessary central e-health information system functionality realization of second part e-health project activities includes integrates e-health information system future development. In e-health project second part according to e-health project development hierarchy plan (Gabaliņa, 2012) are scheduled:

- the existing e-health system future development;
- industry statistics IS development;
- interoperability solutions development for health care institutions;
- information flows support within the European Union (European Patients Smart Open Services (epSOS));
- imaging diagnostics information flow inclusion in e-health;
- financial data processing flow improvement;
- industry standardization continuation.

In the second part is intended to additionally improve the first's part developed information system. There are planned new e-services (LETA, 2013): the ability to authenticate users an additional group, patients self-health input, financial data exchange, GP user interface, industry statistics information system development, connection points to allow the exchange of information between the member states of the European Union.

Third part includes future development of e-health, such as, medical technologies knowledge development, further statistical analysis of the industry, telemedicine e-services etc. Within e-health project third part planned actions (Gabaliņa, 2012) are:

- medical technology knowledge base development;
- further development of the industry statistics IS;
- healthcare welfare of information flow distribution;
- telemedicine e-services;
- medical coding and classification system improvement.

Current status for e-health information system development process

First part of centralized e-health solution has been developed and include e-health integration platform pilot project and e-services, also providing published interfaces for interconnectivity, authentication and classifier distribution services.

In late 2012 were started e-health project testing. Currently are being received results from developers and shortly become to a pilot project where e-health system would operate with real patients data. E-health information system pilot operation (system operation to a limited extent) was scheduled to perform at the end of 2012, but was delayed due to various technical works which was not possible to complete within the time limit, and postponed to 2013 April. Content of the e-health portal is still being completed. There are organized internet survey about e-health portal and what information people would like to see on this portal.

The National Health Service invites hospitals to apply the e-health system testing. Also encouraging medical institutions that have developed their own information systems, as well as medical institutions and other system developers who intend to provide online data transmission between medical institutions information systems and e-health system. Currently contracts are being discussed and signed with medical institutions that have expressed willingness to try out e-health platform to access unified electronic health information system test environment.
After a successful pilot project operation, evaluation of the system will begin - implementation in medical institutions, including GP, outpatient and inpatient care, as well as gradually increase of used functionality.
E-health information systems implementation process is in progress for e-health project 2nd part: the project application is approved, on condition that - the project applicant the National Health Service by March 1 of year 2013 need to provide additional information. NHS has published several informative announcements on planned procurements this year. The procurement specifications are still being prepared. The second part of the project is scheduled to announce procurement for the integrated e-health information systems technical specifications development and for information system development quality assurance. Multiple procurements scheduled also for the first part (the expert services of electronic health records project, e-referrals, e-booking and the project site, as well as e-prescription project). In the second part of e-health project first procurements for information system further development will be held in the middle of this year. (LETA, 2013)
Fully implement e-health programs potential in medical institutions will be possible only after the integration of the central e-health systems. Medical institution or pharmacy may require the integration of e-health systems, implementing some or all of the functions groups: a secure communication channel creation with the central e-health system; records - the exchange of doctor's consultations, services and current appointment information synchronization; prepared prescriptions forwarded to the central e-prescription system, receiving electronic prescriptions (pharmacy) and issued medicines and medical equipment registration; obtaining and maintaining of centralized classifier etc. (VEC e-veselības projektu arhitektūras vadlīnijas, 2010)
For the implementation of e-health center system is required to change the current situation. Currently, hospitals are mainly using few main systems: "*Ārstu birojs*", "*ProfDoc*" and "*Medium*". However in hospitals and other medical institutions there are other solutions to provide effective information systems collaboration between medical institutions and central systems but require changes to existing systems to achieve interoperability.
But thus began the most urgent problems. The health care institutions need to organize e-health implementation in their institution themselves, to seek and ensure funding and so on. If each health care institution implements e-health individually it is very expensive process and not as effective more.

Main e-health implementation problematic

Currently e-health information system provides main role as system for exchange of information between different institutions. In 2013 will conclude the first part of e-health project. First results are expected. In Latvia matters concerning e-health implementation have been discussed in society, by the health sector professionals e-health is ambiguously valuated. Main topics of interest are about:

- Interoperability.

Systems joint and implementation, how it will be done? Who will pay for the existing medical institutions systems modification? Where hospitals and family doctors would get funding? It is not known whether will be available ERDF funding. Problems, that occur while e-health has not been implemented fully and as it is not used completely, it causes complications and expenses. Studies show that at the beginning system causes greater risk of errors. Services for Latvian population might be too expensive, messy legal base for sensitive data protection etc.

- Maintenance.

The Ministry of Health estimates (LETA, 2012) that the annual maintenance would require approximately 700000 EUR. Currently, this funding is intended to cover of the state budget. It is planned that system will pay off after about 6 years. How long will it take for the system to really improve the quality of health care? As mentioned Valdis Keris, The Trade Union of Health and Social Care Employees of Latvia (LVSADA) chairman (Šaboha, 2012): "No mentions are made of the project economic efficiency calculations. This raises concerns about whether such efficiency is generally expected. For example, the implementation of e-prescriptions would mean that any doctor who writes a prescription, the workplace must be equipped with a computer and an internet connection, a computer security purposes to conduct regular maintenance and software upgrades. As far as known, the additional funding for that purpose by the Ministry of Health has not provided."

- Legal and regulatory restrictions (currently many laws impose the need for paper documents with a signature).

The Ministry of Health has developed and submitted for public discussion a draft of the Cabinet of Ministers "Regulations on health information system", aiming to provide an e-health information systems legal basis. But they are very incomplete causing concerns and extensive discussions in society.
In 2009 the Patients' Rights Act was approved. The Act attributes (Šitcs, Giest, Dumortier, Artmann 2010) the patient a right to information, right to medical treatment, right to consent or to abandon of the treatment, the right to free choice and the right to access medical information. The Patients' Rights Act does however not

explicitly mention the use of new e-health techniques such as telemedicine, electronic health records or e-prescription. It can be only assumed that the law is thus equally applicable to paper and electronic records.

- Security and confidentiality of patient's data.

Currently, main topics of the discussion are the problems associated with data security and confidentiality. The patient's medical data reliability and availability implementing the planned e-health information system are widely discussed, as in society has different views and experience on IT ability to ensure data integrity and security. Currently the biggest concerns are about system security, from data leaking, ending with breaking in system and stealing data that is considered sensitive. E-health information system includes a number of security levels and mechanisms (authentication, authorization, data encryption and other internal system security mechanisms) for information systems and data protection. As explains Liene Cipule, Parliamentary Secretary to the Ministry of Health (Cipule L., 2012): "E-health cannot compromise the security of patient data. Data can endanger poorly developed system or a deliberate transgression of the law. The project attracted professionals and expert's assignment is to create a high-quality system and take care of the highest level of safety and protection, developing a tiered and differentiated access to patient data, which will be provided. Industry official's challenge is to adjust to all the rules, regulations and laws that provide e-health activities and movement of data collection and processing procedures. All deficiencies must be corrected before implementation. That is why the draft for public consultation is passed, e-health project scheduled to test mode, a pilot regime and gradual implementation."

These identified risks are related to e-health information system, but should pay attention to the human factor as well, the medical institutions and medical treatment of non-professional patient data security, data disclosure, and legal aspects of the treatment of the patient's sensitive health data - medical records, test results and history collection in corridor or chamber.

Despite all concerns associated to e-health development and implementation process, the benefits of this project is absolutely undeniable and its implementation have to be immediate, although there have been concerns about the potential of e-health system capacity problems and patient data security, etc. These are actual risks of implementing any new large information system project. At the start of the system operation discovered errors have to systemize and avoid, data security have to constantly improve by applying the most advanced protection, also in future perspective managing the system in working order, because of so rapidly change of information technology. Contemporaneity about data security thinks everyone, banks and state institutions, also households, health care is not exception. Society should provide with the most extensive information on e-health use, opportunities and personal data security.

Conclusion

The research shows that in Latvia e-health information system is developed to improve health care efficiency and quality. It will provide centralized storage circulation system of patient health information, providing rapid access within the treatment process involved medical professionals, staff and patients.

The data analysis show that e-health in Latvia is problem issue since 2003. In Latvia e-health implementation has been late because of funding, untenable willingness to accept it etc. Finally in 2013 will conclude e-health project implementation first part. First results are expected with a special attention on the project integration with existing systems and also data security. Currently there is no clarity on e-health project real data security, not always everything planned is reached. Clear is that it is being seriously considered and discussed.

E-health current development process in Latvia is at start of e-health implementation 2nd part process for e-health program and systems development involving sector's administration and professionals, society, IT industry, incorporating best practices across EU.

It is also obviously that e-health implementation is essential not only for the citizens, letting doctors operatively to work, but also it is needed at the national level. Large part of health problems, including financial management solution main key is the e-health. E-health will ensure a transparent health care resources spending, and avoid duplication of investigations, thus providing the economic benefits that will be directed to the patient's treatment.

References

Cipule, L., 2012. LVSADA aicinājums apturēt e-veselību ir bezatbildīgs un neizprotams. Available at: http://www.delfi.lv/news/comment/comment/liene-cipule-lvsada-aicinajums-apturet-e-veselibu-ir-bezatbildigs-un-neizprotams.d?id=42577700, 15.12.2012.

Gabaliņa, D., 2012. E-health in Latvia, The Ministry of Health, Latvia

Gartner Group, 2009. eHealth for a Healthier Europe!, Sweden, 17 p.

LETA, 2012. E-veselības pilotdarbināšana atlikta uz nākamo gadu. Available at: http://www.apollo.lv/zinas/e-veselibas-pilotdarbinasana-atlikta-uz-nakamo-gadu/542327. 10.01.2013. (In Latvian)

LETA, 2013. E-veselības otrās kārtas projekts apstiprināts. Available at: http://www.focus.lv/content/e-veselibas-otras-kartas-projekts-apstiprinats, 02.02.2013. (In Latvian)

Nacionālais veselības dienests, 2012. Darbības programmas "Infrastruktūra un pakalpojumi" papildinājuma 3.2.2.1.1.apakšaktivitātes „Informācijas sistēmu un elektronisko pakalpojumu attīstība" projekta „E-veselības integrētās informācijas sistēmas attīstība" e-veselības informācijas sistēmas darbības koncepcijas apraksts. Rīga, pp. 5-58. (In Latvian)

Ričika, D., Zonne, I., 2006. Digitālās slimības vēstures, receptes un citi brīnumi. Available at: http://www.doctus.lv/2006/10/digitalas-slimibas-vestures-receptes-un-citi-brinumi, 15.01.2013. (In Latvian)

Springe, R., 2011. E-health framework implementation, IS Architecture concepts and solutions in context of Estonian experience and Latvia's challenge, Centre of Health Economics, Latvia

Šaboha, I., 2012. Pacientu datu drošums e-veselībā. LV portālam: VALDIS KERIS, Latvijas Veselības un sociālās aprūpes darbinieku arodbiedrības (LVSADA) priekšsēdētājs. Available at: http://www.lvportals.lv/viedokli.php?id=250398, 15.12.2012. (In Latvian)

Šitcs, M., Giest, S., Dumortier, J., Artmann, J., 2010, Country Brief: Latvia, European Commission, DG Information Society and Media, ICT for Health Unit, 30 p.

Technology Development Forum, 2011. Analytical reports for focus application areas LATVIA – eHealth, O3.8,pp. 2-11.

The National Health Service Latvia, 2012. E-health Latvia. Available at: http://www.vmnvd.gov.lv/en/e-health, 25.01.2013.

Veselības Ekonomikas centrs, 2010, VEC e-veselības projektu arhitektūras vadlīnijas, Latvia. pp. 16-30. (In Latvian)

WSIS STOCKTAKING & PARTNERSHIPS : ACTIVITY DETAILS, 2013. "E-health in Latvia" - development and implementation of national e-health policy. Available at: http://www.itu.int/wsis/stocktaking/scripts/documents.asp?project=1270136996, 04.01.2013.

Zvejniece, R., 2011. Veselības dati – no papīra uz elektronisku informācijas apriti, LV portālam: DAIGA BEHMANE, Veselības ekonomikas centra direktore. Available at: http://lvportals.lv/viedokli.php?id=232743, 05.12.2012. (In Latvian)

Smart Cities Software from the developer's point of view

Dmitry Namiot[1], Manfred Schneps-Schneppe[2]

[1]*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia dnamiot@gmail.com*
[2]*Ventspils International Radio Astronomy Centre Ventspils University College, Ventspils, Latvia manfreds.sneps@gmail.com*

Abstract: *The paper discusses the current state and development proposals for Smart Cities and Future Internet projects. Definitions of a Smart City can vary but usually tend to suggest the use of innovative Info-Communication technologies such as the Internet of Things and Web 2.0 to deliver more effective and efficient public services that improve living and working conditions and create more sustainable urban environments. Our goal is to analyze the current proposals from the developer's point of view, highlight the really new elements, the positions borrowed from the existing tools as well as propose some new extensions. We would like to discuss the possible extensions for the existing proposals and describe add-ons that, by our opinion, let keep the future research inline with the modern approaches in the web development domain.*

Keywords: m2m, future internet, open API, Smart City, Web Intents.

Introduction

The concept of "smart cities" has attracted considerable attention in the context of urban development policies. The Internet and broadband network technologies as enablers of e-services become more and more important for urban development while cities are increasingly assuming a critical role as drivers of innovation in areas such as health, inclusion, environment and business (Schaffers et al., 2011). There is a limitless amount of available online resource and articles, devoted to smart cities and related IT tools. Our goal in this article is to analyze them from the software development point of view.

There are actually more than one defintion for the term „smart city". For example, IBM's Smart Cities digitize and connect infrastructures (IOT) to infuse them with new intelligence (IBM, 2012). As per Forrester, Smart City is the combined use of software systems, server infrastructure, network infrastructure, and client devices to better connect seven critical city infrastructure components and services: city administration, education, healthcare, public safety, real estate, transportation, and utilities (Smart City, 2012).

The rest of the paper is organized as follows. We consider two projects: EPIC project, and FI-Ware project. Then we analyze the Open API for M2M, submitted to ETSI, and offer a new web tool, Web Intents for M2M, as an enhancement of M2M middleware.

EPIC project

Altogether, seven pilot projects were selected for the call for funding under the CIP ICT Policy Support Program on "Open Innovation for Future Internet-enabled services in smart cities". This list includes the following projects: EPIC, Smart IP, Peripheria, People, Open Cities, Life 2.0 and Smart-Islands.

The European Platform for Intelligent Cities (EPIC) is a European Commission-funded project that aims to wed state-of-the-art cloud computing technologies with fully researched and tested e-Government service applications to create the first truly scalable and flexible pan-European platform for innovative, user-driven public service delivery.

The EPIC platform will combine the industrial strength of IBM's 'Smart City' vision and cloud computing infrastructure with the knowledge and expertise of the Living Lab approach (which expressly engages citizens in service design) to ensure the development of a European 'innovation ecosystem' to deliver sustainable, user-driven web-based services for citizens and businesses.

The EPIC project will help to significantly accelerate the uptake of new citizen-generated services across Europe by combining the world-leading business expertise of Deloitte Consulting with the practical, first-hand knowledge of the European Network of Living Labs to guide cities through the routes, decisions and steps they need to undertake to improve service delivery and achieve the benefits of 'smart' working. Fig. 1 presents the whole project.

The key points, from the development point of view, are: test and development cloud and its APIs, service catalogue and its API's (at least discovery and publishing), and IoT middleware.

At this moment, we can already highlight one significant moment: the word "middleware" actually describes the very wide area of software. And it is not understandable, for example, why cloud and directory of services are separated from the middleware? All three components could be presented actually as middleware.

In the terms of software, EPIC approach is based on the well known Web services approach in the traditional (canonical) form: WSDL, SOAP, etc. At this stage, EPIC does not introduce really new elements for the developers community.

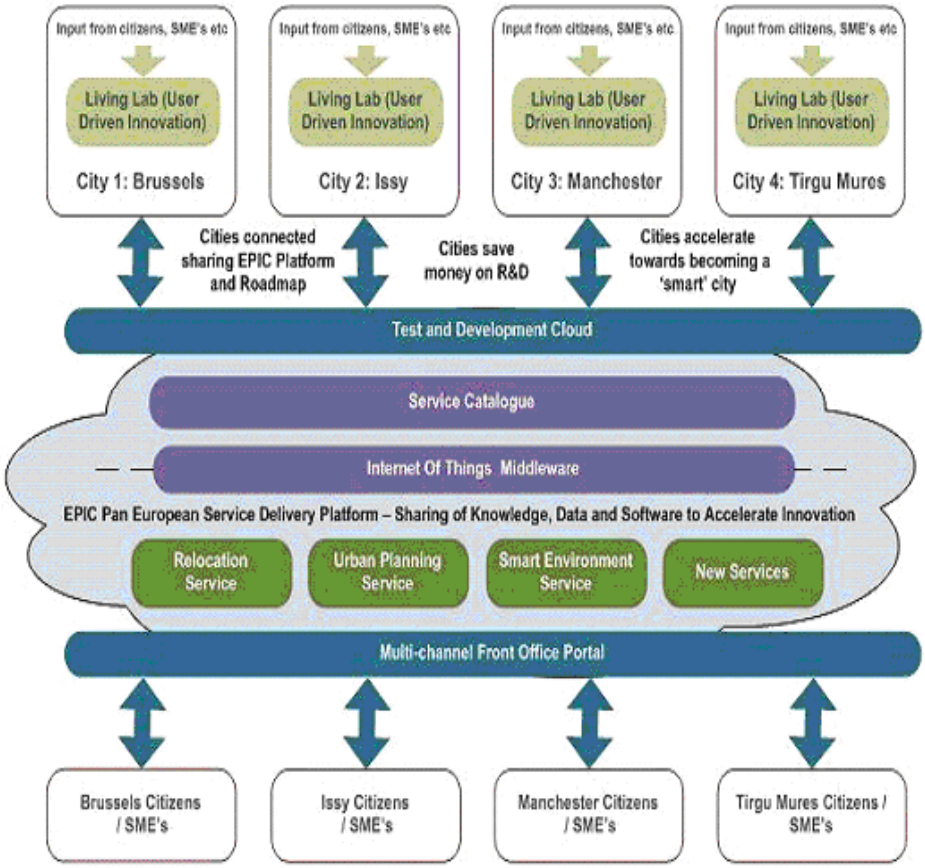


Fig. 1**. EPIC project roadmap** (EPIC, 2012).

FI-Ware project

The more interesting from the developer's point of view is FI-Ware project. FI-WARE is being developed as part of the Future Internet Public Private Partnership program launched by the European Commission in collaboration with the ICT Industry.

The Reference Architecture of the FI-WARE platform is structured along a number of technical chapters, namely: Cloud Hosting, Data/Context Management, Internet of Things (IoT) Services Enablement, Applications/Services Ecosystem and Delivery Framework, Security and Interface to Networks and Devices. As per official document, FI-WARE will enable smarter, more customized/personalized and context-aware applications and services by the means of a set of assets able to gather, exchange, process and analyze massive data in a fast and efficient way (Fig. 2).

Data in FI-WARE refers to information that is produced, generated, collected or observed that may be relevant for processing, carrying out further analysis and knowledge extraction. A basic concept in FI-WARE is that data elements are not bound to a specific format representation. Actually the whole data model in FI-WARE has been described with the concept of NoSql systems in mind (Pokorny, 2011). Data items could be named and presents themselves by just named collection of triples: <name, type, value>.

What is important, that optionally, data elements could have meta-data (descriptions) associated with them. Meta-data elements could be described via collections of triples <name, type, value> too. The data-model described in FI-WARE is actually could be perfectly supported by distributed key-value systems (Lakshman, Malik, 2010).

Context in FI-WARE is represented through context elements. A context element extends the concept of data element by associating an *EntityId* and *EntityType* to it, uniquely identifying the entity (which in turn may map to a group of entities) in the FI-WARE system to which the context element information refers. In addition, there may be some attributes as well as meta-data associated to attributes that we may define as mandatory for context elements as compared to data elements (FI-Ware, 2012b).

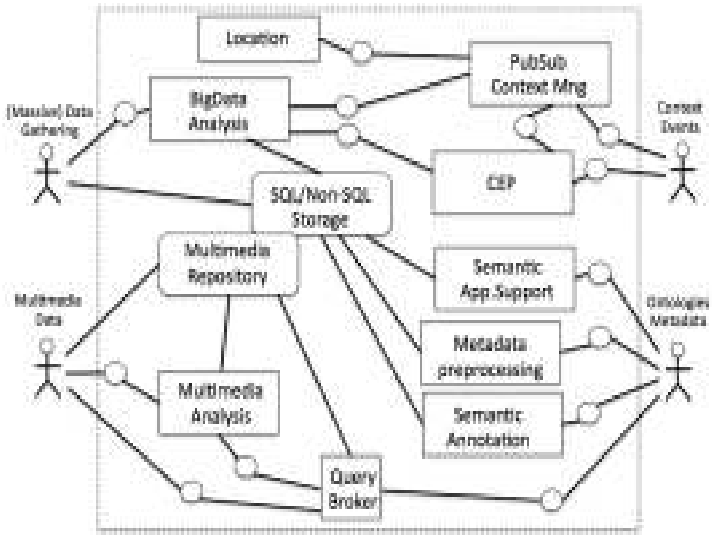


Fig. 2. **FI-WARE project** (FI-Ware, 2012a).

It is very important that FI-WARE actually uses the same model for data and meta-data. It means that from the developer's point of view it should be possible to use the same model for persistence and search for both data and meta-data.

An event in FI-WARE is an act of creation a new element. It could be either data event (create data elements), or context event (creates context element). As an example, a sensor device measures some value and periodically creates and sends a new context element. The creation and sending of the context element is an event.

Because each event has got either data or context elements linked to, the whole system can see events via linked data. It makes the whole system much more uniform (homogeneous) comparing with M2M approach described below.

Discussion

At this moment we have a wide choice for real-time analytical systems based on key-value stores. For example, we can mention Google Percolator (Peng and Dabek, 2010) or Twitter Storm (Marz, 2011). It is exactly the approach needed for processing data in FI-WARE model. For event publishing, FI-WARE roadmap suggests ContextML (Knappmeyer et al., 2010) and SPARQL (Sparql, 2012).

ContextML is light-weight XML based context representation schema in which context information is categorized into scopes and related to different types of entities (e.g. user, device). The schema is also applied for encoding management messages in order to allow for a flexible framework supporting gradual plug & play extendibility and mobility. *ContextML* is tailored to be used for REST-based communication between the framework components.

SPARQL. RDF is a directed, labeled graph data format for representing information in the Web. And SPARQL specification defines the syntax and semantics of the query language for RDF. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs.

O&M example. The obvious candidates here are standards activities from The Open Geospatial Consortium (OGC) that focus on sensors and sensor networks comprise (Botts, 2008). At the first hand it is Observations & Measurements Schema (O&M) as well as Sensor Model Language (SensorML), Transducer Model Language (TransducerML or TML), Sensor Observations Service (SOS), Sensor Planning Service (SPS) and Sensor Alert Service (SAS).

For example O&M supports data sampling as this:

```
<gml:description>
   Observation test instance: fruit mass
</gml:description>
 <gml:name>Observation test 1</gml:name>
  <om:phenomenonTime>
 <gml:TimeInstant    gml:id="ot1t">
<gml:timePosition>
2005-01-11T16:22:25.00
```

```
</gml:timePosition>
   </gml:TimeInstant>
 </om:phenomenonTime>
  <om:parameter>
<om:NamedValue>
<om:name xlink:href="http://sweet.jpl.nasa.gov/ontology/property.owl#Temper
ature"/>
 <om:value xsi:type="gml:MeasureType" uom="Cel">
22.3
</om:value>
</om:NamedValue>
</om:parameter>
```

XML vs. JSON. But keeping in mind the modern trend in web development – shall we keep that as XML, or it is a time to replace it with an appropriate JSON?

FI-WARE proposes also an interesting approach for Applications/Services Ecosystem and Delivery Framework. It is based on the heavy usage on USDL (Barros, 2012). Universal Service-Semantics Description Language (USDL) can be used by service developers to specify formal semantics of web-services. Thus, if WSDL can be regarded as a language for formally specifying the syntax of web services, USDL can be regarded as a language for formally specifying their semantics. USDL is as formal service documentation that will allow sophisticated conceptual modeling and searching of available web-services, automated composition, and other forms of automated service integration. For example, the WSDL syntax and USDL semantics of web services can be published in a directory which applications can access to automatically discover services (Park, 2011).

Open API for M2M

Why we think that it is a must to talk about M2M API here? Actually, in the many aspects, IoT and M2M applications are performing absolutely the same tasks. For example, both systems could be described as data collectors for the future processing. This section describes an Open API for M2M, submitted to ETSI. It is probably the most valuable achievement at this moment. Roadmap for FI-WARE highlights the plans for M2M General Enabler (Gateway). In the same time M2M Open API can benefits from more elaborated FI-WARE persistence and cloud offerings.

The Open API for M2M applications are developed jointly in EURESCOM study P1957 and the EU FP7 SENSEI project. The Open API has been submitted as a contribution to ETSI TC M2M for standardization (ETSI, 2011).

Actually, in this Open API, we can see the big influence of Parlay specification. Parlay Group leads the standard, so called Parlay/OSA API, to open up the networks by defining, establishing, and supporting a common industry-standard APIs. Parlay Group also specifies the Parlay Web services API, also known as Parlay X API, which is much simpler than Parlay/OSA API to enable IT developers to use it without network expertise (Yim, 2006).

The goals are obvious, and they are probably the same as for any unified API. One of the main challenges in order to support easy development of M2M services and applications will be to make M2M network protocols "transparent" to applications. Providing standard interfaces to service and application providers in a network independent way will allow service portability (Gronbek, 2008).

At the same time, an application could provide services via different M2M networks using different technologies as long as the same API is supported and used. By this way, an API shields applications from the underlying technologies and reduces efforts involved in service development. Services may be replicated and ported between different execution environments and hardware platforms (Gronbek, Ostendorf, 2010).

This approach also lets services and technology platforms to evolve independently. A standard open M2M API with network support will ensure service interoperability and allow ubiquitous end-to-end service provisioning.

The Open API provides service capabilities that are to be shared by different applications. Service Capabilities may be M2M specific or generic, i.e., providing support to more than one M2M application.

Key points for Open API are:

- it supports interoperability across heterogeneous transports
- ETSI describes high-level flow and does not dictate implementation technology
- it is message-based solution
- it combines P2P with client-server model
- and it supports routing via intermediaries

At this moment, all point are probably not discussable except the message-based decision. Nowadays, publish-subscribe method is definitely not among the favorites approaches in the web development, especially for heavy-loading projects.

Let us name the main Open API categories (Table 1) and make some remarks (Sneps-Sneppe, Namiot, 2012a).

Table 1

Open API categories and remarks

Open API categories	**API contents**	**Comments**
Grouping	A group here is defined as a common set of attributes (data elements) shared between member elements. On practice it is about the definition of addressable and exchangeable data sets.	Just note, as it is important for our future suggestions, there are no persistence mechanisms for groups.
Transactions	Service capability features and their service primitives optionally include a transaction ID in order to allow relevant service capabilities to be part of a transaction. Just for the deploying transactions and presenting some sequences of operations as atomic.	In the terms of transactions management Open API presents the classical 2-phase commit model. By the way, we should note here that this model practically does not work in the large-scale web applications. We think it is very important because without scalability we cannot think about "billions of connected devices".
Application Interaction	The application interaction part is added in order to support development of simple M2M applications with only minor application specific data definitions: readings, observations and commands.	Application interactions build on the generic messaging and transaction functionality and offer capabilities considered sufficient for most simple application domains.
Messaging	The Message service capability feature offers message delivery with no message duplication. Messages may be unconfirmed, confirmed or transaction controlled.	The message modes supported are single Object messaging, Object group messaging, and any object messaging; (it can also be Selective object messaging). Think about this as Message Broker.
Event notification and presence	The notification service capability feature is more generic than handling only presence. It could give notifications on an object entering or leaving a specific group, reaching a certain location area, sensor readings outside a predefined band, an alarm, etc.	It is a generic form. So, for example, geo fencing should fall into this category too. The subscriber subscribes for events happening at the Target at a Registrar. The Registrar and the Target might be the same object. This configuration offers a publish/subscribe mechanism with no central point of failure.
Compensation	Fair and flexible compensation schemes between cooperating and competing parties are required to correlate resource consumption and cost, e.g. in order to avoid anomalous resource consumption and blocking of incentives for investments. The defined capability feature for micro-payment additionally allows charging for consumed network resources.	It is very similar, by the way, to Parlay's offering for Charging API.
Sessions	In the context of OpenAPI a session shall be understood to represent the state of active communication between Connected Objects.	OpenAPI is REST based, so, the endpoints should be presented as some URI's capable to accept (in this implementation) the basic commands GET, POST, PUT, DELETE (See an example below).

A session example: requests execution of some function.
URI: http://{nodeId}/a/do

```
Method: POST
Request
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<appint-do-request xmlns="http://eurescom.eu/p1957/openm2m">
<requestor>9378f697-773e-4c8b-8c89-27d45ecc70c7</requestor>
<commands>
```

```
<command>command1</command>
<command>command2</command>
</commands>
<responders>9870f7b6-bc47-47df-b670-2227ac5aaa2d</responders>
<transaction-id>AEDF7D2C67BB4C7DB7615856868057C3</transaction-id>
</appint-do-request>
Response
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<appint-do-response xmlns="http://eurescom.eu/p1957/openm2m">
<requestor>9378f697-773e-4c8b-8c89-27d45ecc70c7</requestor>
<timestamp>2010-04-30T14:12:34.796+02:00</timestamp>
<responders>9870f7b6-bc47-47df-b670-2227ac5aaa2d</responders>
<result>200</result>
</appint-do-response>
```

Note that because we are talking about server-side solution, there is no problem with so called sandbox restrictions. But it means of course, that such kind of request could not be provided right from the client side as many modern web applications do. In the same time FI-WARE authors, at least according to roadmap, pays attention to JSON and other client-side technologies. The reasons are obvious – the modern web-based technologies can save the development time. The next section below is also devoted to client-side technologies.

Data persistence. Another area, where M2M API in the current form is weak by our opinion, belongs to data persistence. We should keep in mind that we are talking about the particular domain – M2M. In the most cases, all business applications here will deal with some metering data. As soon as we admit, that we are dealing with the measurements in the various forms, we should make, as seems to us a natural conclusion – we need to save the data somewhere. It is the core business for M2M – save data for the future processing.

So, the question is very easy – can we talk about M2M applications without talking about data persistence? Again, the key question is M2M. It is not some abstract web API. We talk about the well-defined domain here.

As seems to us, even right now, before the putting some unified API in place, the term M2M almost always coexists with the term "cloud". And as we can see, almost always has been accompanied by the terms like automatic database logging, backup capabilities, etc.

So, maybe this question is more for the discussions or it even could be provocative in the some forms, but it is: why there is no reference API for persistence layer in the unified M2M API? It is possible in general to create data gathering API without even mentioning data persistence?

Web intents vs. Open API from ETSI

This section is devoted to the relatively new approach in the client-side web development – Web Intents. The first time Web Intents usage for M2M applications was proposed by Dmitry Namiot (Sneps-Sneppe, Namiot, 2012b).

Let us start from the basic. Users use many different services on the web to handle their day to day tasks, developers use different services for various tasks. In other words, our environment consists of connected applications. And of course, all they expect their applications to be connected and to work together seamlessly.

It is almost impossible for developers to anticipate every new service and to integrate with every existing external service that their users prefer, and thus, they must choose to integrate with a few select APIs at great expense to the developer.

As per telecom experience, we can mention here the various attempts for unified API that started, probably, with Parlay. Despite a lot of efforts, Parlay API's actually increase the time for development. It is, by our opinion, the main reason for the Parlay's failure.

Web Intents solves this problem. Web Intents is a framework for client-side service discovery and inter-application communication. Services register their intention to be able to handle an action on the user's behalf. Applications request to start an action of a certain verb (for example: share, edit, view, pick, etc.) and the system will find the appropriate services for the user to use based on the user's preference. It is the basic (Namiot, 2013).

Going to M2M applications it means that our potential devices will be able to present more integrated for the measurement visualization for example. The final goal of any M2M based application is to get (collect) measurements and perform some calculations (make some decisions) on the collected dataset. We can go either via low level API's or use (at least for majority of use cases) some integrated solutions. The advantages are obvious. We can seriously decrease the time for development.

Web Intents example. Web intents put the user in control of service integrations and make the developers life simple. Here is the modified example for web intents integration for the hypothetical web intents example:

1. Register some intent upon loading our HTML document

```
document.addEventListener("DOMContentLoaded", function() {
      var regBtn = document.getElementById("register");
      regBtn.addEventListener("click", function() {
```

```
      window.navigator.register("http://webintents.org/m2m", undefined);
      }, false);
2. Start intent's activity
      var startButton = document.getElementById("startActivity");
      startButton.addEventListener("click", function() {
        var intent = new Intent();
        intent.action = "http://webintents.org/m2m";
          window.navigator.startActivity(intent);
      }, false);
3. Get measurements (note - in JSON rather than XML) and display them in
our application
      window.navigator.onActivity = function(data) {
        var output = document.getElementById("output");
        output.textContent = JSON.stringify(data);
      };
    }, false);
```

Discussion. Obviously, that it is much shorter than the long sequence of individual calls as per M2M Open API. The key point here is *onActivity* callback that returns JSON (not XML!) formatted data. As per suggested M2M API we should perform several individual requests, parse XML responses for the each of them and only after that make some visualization. Additionally, web intents based approach is asynchronous by its nature, so, we do not need to organize asynchronous calls by our own.

Also, Web Intents approach lets us bypass sandbox restrictions. In other words, developers can raise requests right from the end-user devices, rather than always call the server. The server-side only solution becomes bottleneck very fast. And vice-versa, client side based requests let developers deploy new services very quickly. Why do not use the powerful browsers in the modern smart-phones? At the end of the day, Parlay specifications were born in the time of WAP and very rudimentary mobile browsers. Why do we ignore HTML5 browsers and JavaScript support in the modern phones?

Future work

Considering M2M communications as a central point of Future Internet, European commission creates standardization mandate M/441. The Standardization mandate M/441, issued on 12th March 2009 by the European mandate to CEN, CENELEC and ETSI, in the field of measuring instruments for the development of an open architecture for utility meters involving communication protocols enabling interoperability, is a major development in shaping the future European standards for smart metering and Advanced Metering Infrastructures. The general objective of the mandate is to ensure European standards that will enable interoperability of utility meters (water, gas, electricity, heat), which can then improve the means by which customers' awareness of actual consumption can be raised in order to allow timely adaptation to their demands.

ETSI's 3rd workshop on Machine to Machine (M2M) communications, held in Mandelieu, France on October 23-25, 2012, gathered leading experts from all over the world to hear how ETSI M2M technology standards are being deployed. "With 270 registered delegates from four continents, 25 speakers, thirteen live demonstrations of M2M-based applications and two days of intense discussion, this year's event was again a success," – the official workshops' site assure us (M2MWORKSHOP, 2012). In reality, the state-of-the-art with M2M standards is far from hopeful.

The demos covered once again a respectable cross section of the application domains such as: Smart Metering, Home automation, Energy Efficiency, Smart building, Smart City, Smart Parking, Exercise, Gaming and Home Energy, Management Systems linked with Social Networking Service and others. But the existing standards CoAP, 6lowpan, ETSI M2M, OMA DM, BBF TR069, OSGi, HGI, etc. and the ZigBee, KNX, etc. are far from convergence.

The new international M2M Partnership Project "oneM2M" is now started. The list of funding partners include ETSI (Europe), ATIS and TIA (US), CCSA (China), TTA (Korea), ARIB and TTC (Japan). And the leading role of ETSI goes more sophisticated.

Conclusions

This paper discusses the current state and development proposals for Smart Cities and Future Internet projects. Our goal was to analyze the current proposals from the developer's point of view. Software development companies and universities should start investigate tools used and proposed in the Future Internet roadmap right now. Also we discussed the possible extensions for the existing proposals and describe the add-ons that, by our opinion, let keep the future research inline with the modern approaches in the web development domain. Article proposes some new additions – web intents as add-on for the more traditional REST approach. The main goal for our suggestions is the simplifying of the development phases for new applications by support asynchronous calls and JSON versus XML data transfer.

Acknowledgments

The paper is financed from ERDF's project SATTEH (No. 2010/0189/2DP/2.1.1.2.0/10/APIA/VIAA/019) being implemented in Engineering Research Institute «Ventspils International Radio Astronomy Centre» of Ventspils University College (VIRAC).

References

Barros, A., Oberle, D., Kylau, U., Heinzl, S., 2012. *Handbook of Service Description. Design Overview of USDL.* Part 2, pp.187-225.

Botts, M., Percivall, G., Reed, C., Davidson, J., 2008. OGC Sensor Web Enablement: Overview and High Level Architecture. *GeoSensor Networks. Lecture Notes in Computer Science,* Volume 4540/2008, pp.175-190.

EPIC, 2012. Available at: http://www.epic-cities.eu/ Retrieved: Jan, 2013.

ETSI, 2011. Draft ETSI TS 102 690 V0.13.3 (2011-07) Technical Specification.

FI-Ware, 2012a. Available at: http://www.finest-ppp.eu/files/deliverables/d03/finest_d3.2.pdf Retrieved: Feb, 2013.

FI-Ware, 2012b. Available at: http://forge.fi-ware.eu/plugins/mediawiki/wiki/fiware/index.php Retrieved: Jan, 2013.

Grønbæk, I., 2008. Architecture for the Internet of Things (IoT): API and interconnect, *The Second International Conference on Sensor Technologies and Applications,* IEEE August 2008.

Grønbæk, I. and Ostendorf, K., 2010. Open API for M2M applications. *ETSI M2M Workshop,* Oct. 2010.

IBM, 2012. Available at: http://www.ibm.com/smarterplanet/uk/en/sustainable_cities/ideas/index.html Retrieved: Feb, 2013

Knappmeyer, M., Kiani, S.L., Fra, C., Moltchanov, B., Baker, N., 2010. ContextML: A light-weight context representation and context management schema. *Wireless Pervasive Computing (ISWPC), 5th IEEE International Symposium on Date of Conference:* 5-7 May 2010, pp. 367 – 372.

Lakshman, A. and Malik, P., 2010. Cassandra: a decentralized structured storage system. *ACM SIGOPS Operating Systems Review archive,* Volume 44. Issue 2, April 2010, pp. 35-40.

M2MWORKSHOP, 2012. Available at: http://www.etsi.org/ Retrieved: Feb, 2013

Marz, N., 2011. A Storm is coming. Available at: http://engineering.twitter.com/2011/08/storm-is-coming-more-details-and-plans.html, August 2011.

Namiot, D., 2013.. "Web Intents as an Extension for M2M Open API." European Journal of Scientific Research 94.3 (2013): pp. 452-462.

Park, H., Kim, B., Ko, Y., Lee, D., 2011. InterX: A service interoperability gateway for heterogeneous smart objects. *Pervasive Computing and Communications Workshops (PERCOM Workshops), IEEE International Conference,* 21-25 March 2011, pp. 233 – 238.

Peng, D. and Dabek, F., 2010. Large-scale incremental processing using distributed transactions and notifications. *OSDI '10*, pp. 251–264.

Pokorny, J., 2011. NoSQL databases: a step to database scalability in web environment. *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services,* ACM, New York, USA, pp. 278-283.

Schaffers, H. et al., 2011. Smart Cities and the Future Internet: Towards Cooperation Frameworks for Innovation. *J. Domingue et al. (Eds.), The Future Internet,* Springer, pp. 431-446.

Schneps-Schneppe, M., Maximenko, A., Namiot, D., Malov, D., 2012. Wired Smart Home: energy metering, security, and emergency issues. In Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on (pp. 405-410). IEEE.

SPARQL, 2012. Available at: http://www.mendeley.com/research/sparql-query-language-for-rdf/ Retrieved: Feb, 2013.

Sneps-Sneppe, M. and Namiot, D. (2012, April). About M2M standards. M2M and Open API. In ICDT 2012, The Seventh International Conference on Digital Telecommunications (pp. 111-117).

Sneps-Sneppe, M., and Namiot, D., 2012. "M2M Applications and Open API: What Could Be Next?." Internet of Things, Smart Spaces, and Next Generation Networking (2012): pp. 429-439..

Yim, J., Choi, Y., Lee, B., 2006. Third Party Call Control in IMS using Parlay Web Service Gateway. *Advanced Communication Technology, ICACT 2006. The 8th International Conference.* 20-22 Feb. 2006, pp. 221 – 224.

Structure and technical solutions for Web-conferencing system

Alexei Machovikov[1,2]

[1]*Department of Informatics and Computer Technologies, National University of Mineral Resources 'Mining', 2, 21st Line, St-Petersburg, Russia, 199106*
[2]*Department of Software Development, Bradon Technologies Ltd., 5225 Orbitor Drive, Unit 22, Mississauga, Ontario, Canada, L4W 4Y8*
amachovikov@gmail.com

Abstract: *In this paper the structure of web-conferencing system named SAViiMeeting is described. This system contains server part which consists from web-portal, database, database-proxy, media-server, http-proxy, sip-gate and mobile-proxy and clients for Microsoft Windows, Microsoft Windows Phone, BlackBerry and Android. Also some technical solutions which have been developed for this system such as method of heterogeneous data synchronization, frame independent audio codecs, method of real-time data transfer through HTTP and UDP-based reliable data transfer protocol are considered. SAViiMeeting's audio codecs have been noted in 2009 Unified Communications TMC Labs Innovation Award Winners and SAViiMeeting system in whole has been named as Unified Communications Product of the Year in 2008 and 2009.*

Keywords: web-conferencing systems, frame independent audio codecs, real-time data transfer protocols.

Introduction

Web-conferencing systems such as Cisco WebEx and Adobe Connect Pro are increasingly used now for production management and distance learning. One such system has been developed by Bradon Technologies Ltd. (Canada) in 2006-2010. This system was called SAViiDesk and later renamed to SAViiMeeting ('SAVii' is abbreviation for Synchronized Audio Video Interactivity through Internet which is consonant with the word 'savvy'). In this paper I would like to present the system structure and some technical solutions which have been designed by me.

System Structure

Like the overwhelming majority of Internet network services, SAViiMeeting system uses client/server technology.
The client end of the system is represented by applications for Microsoft Windows and Microsoft Windows Phone. There are also lightweight Java-versions of clients for mobile devices of BlackBerry and Android types. Landline, mobile and IP-telephones can be used as 'only-audio' clients.
The server end of the system is represented by a combination of servers, proxy and gateways operating under the control of Linux or Microsoft Windows. It includes:

- Web-portal of SAViiMeeting system which is responsible for user authorization, conference planning, distribution of invitations to conference participants, run client applications, etc.
- Database management system (DBMS) used for storage of information about SAViiMeeting clients, conferences held by them, etc. The fifth version of system can use MySQL or DB2.
- Database proxy is used for creation of an encrypted channel for communication between DBMS and other components of the server end. Such organization allows installation of the transmitting components at the customer's site.
- Media-server is the main transmitting component of the system. Clients connect to the Media-server and send data through it. Interaction of clients and other components of the server end with the Media-server is realized with using of secure UDP-based protocol. The task of the Media-server also is archiving of conferences and reproducing them at the clients' requests.
- HTTP-proxy that used to connect clients that failed to connect to the Media-server directly. The main reason of connection errors is caused by presence of a firewall which blocks the UDP protocol. The client operates with the HTTP-proxy by the HTTP protocol which is allowed by all firewalls.
- SIP-gate which is a gateway to IP-telephony provider. IP-telephony provider receives calls from 'only-audio' clients that are connected to it. If an IP-telephony client is based on SIP/RTP-protocols, it can connect to the SIP-gate directly.
- Mobile-proxy which serves light Java-versions of clients for mobile devices.

The fifth version of SAViiMeeting system can serve three types of customers: public (individual), small and medium business (SMB) and enterprise. The information flows in the SAViiMeeting system are shown in Fig. 1.

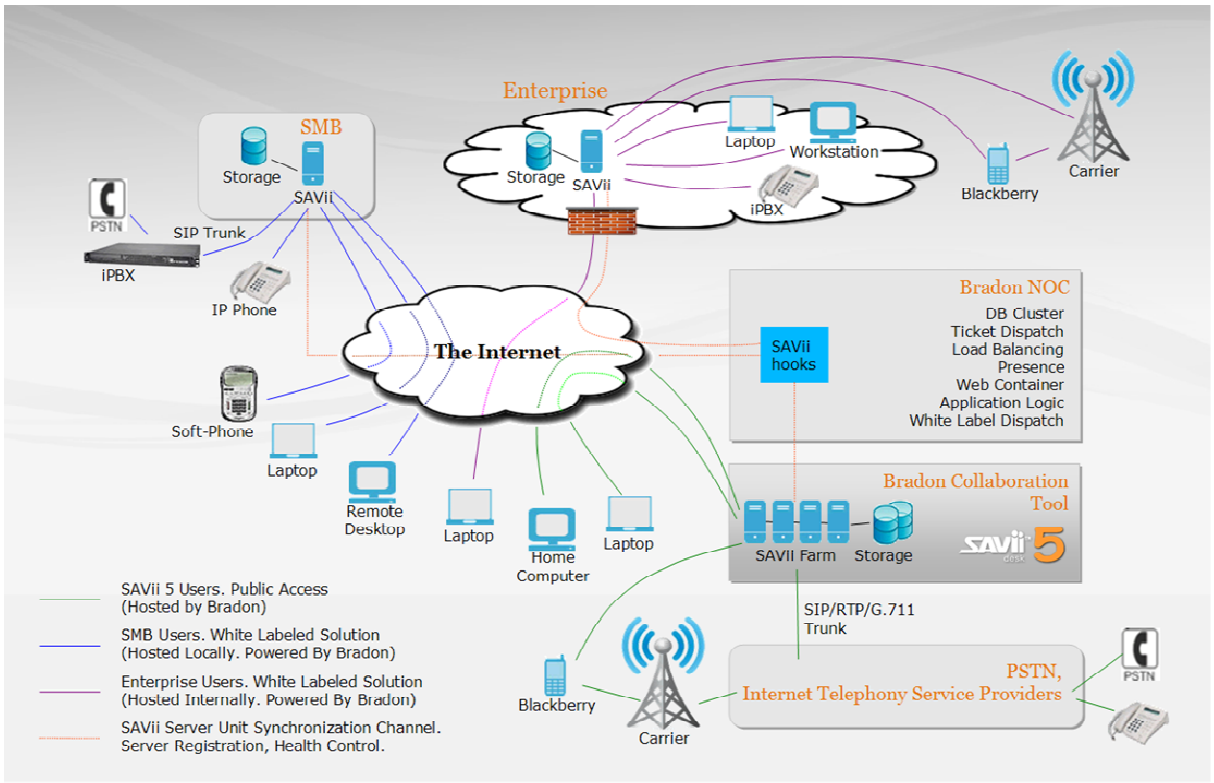


Fig. 1. **The information flows in SAViiMeeting system**.

The work with SAViiMeeting system can be described in the following way. The conference host enters the Web-portal with using of the name and password received upon registration. Using tools implemented in the Web-portal, the host plans the conference by choosing the time for it and invites participants. New participants are added by e-mail addresses. Each participant receives an invitation e-mail with the link to run the client application and the telephone number with the pin code for connecting via the telephone. At the appointed time the host runs the client application for Microsoft Windows and waits for participants to connect. Participants also run its clients and enter the conference. Those who have a computer with a soundcard, a microphone and speakers/earphones run a full client. Later the host can transfer his rights to such participants. Others either run a client in the 'deaf-and-mute' mode and use the telephone or a mobile device. The client for Microsoft Windows Phone is fully functional. Java-clients for BlackBerry and Android can only receive images from the host's desktop. After a sufficient number of participants have connected to the conference, the host starts the presentation. He turns on his microphone and a video-camera, places presentation materials on his desktop, marks them with a frame and enables screen capture. If required, he has a possibility of drawing with a marker inside the captured field. Participants viewing the host's presentation and can speak with him and other participants if their microphones are not blocked. The participant can attract the host's attention by 'raising his hand' (pushing a special button). If the host considers it necessary, he can transfer the right of leading the presentation to one of the participants. The conference can be archived for further viewing.
The desktop of host's computer with SAViiMeeting application interface is shown in Fig. 2.

Some technical solutions

During development of SAViiMeeting systems some serious technical problems have been solved. I would like to describe four such problems and their solutions.
Problem 1. Synchronization of heterogeneous data.
In the web-conferencing system there is a problem of synchronous playback of heterogeneous data such as speech of the host, its video image, images from his desktop, etc. To solve this problem a lot of different solutions were proposed. First, many try to pass all the data in a single stream and create a tight synchronization between heterogeneous data. However, this method of transmission has serious problems with interruption of the data flow in case of bad communication channel. These interruptions the stronger, the greater the amount of data transmitted per unit of time. So, the bulk data which do not require very stronger continuity (for example, images from the desktop) brings a negative effect to the transfer of data that require it, but take a small amount in the stream (for example, speech). Another more correct way is to pass different data in different flows. In this case synchronization between data streams is organized, as a rule, with the using of time stamps. This method is realized, for example, in RTP (Real-time Transport Protocol). But the transport layer there is no information

about the physical meaning of transmitted data, so data which require continuity (speech) can be stopped before the arrival of the data which do not require continuity (image from the desktop).

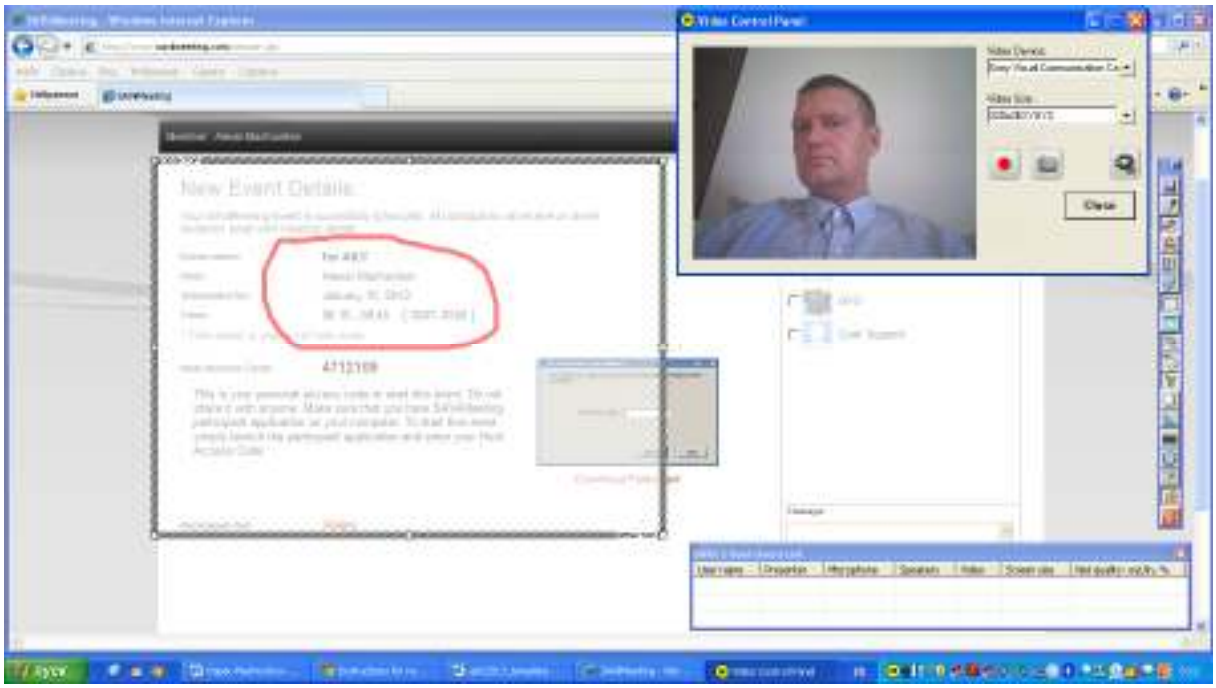


Fig. 2. **Desktop of host's computer**.

So, I came to the idea of 'natural' data synchronization. Each data stream I try to transfer and play in real time. The data which are late are discarded. Request for the retransmission of the lost and corrupted data is not done. Accordingly, all the data are played simultaneously. Many years of successful operation of the SAViiMeeting system is confirmed the validity of this idea.

Problem 2. Audio compression for transmission over a channel with losses.

In 1999-2002, we with Kirill Stolyarov are developed some low speed frame independent audio codecs for 2400, 4800, 9600 and 19200 bps (Machovikov et al., 2001) which can be used in communication channels with losses without additional methods of data protection, such as FEC (Forward Error Correction). These codecs are used in SAViiMeeting system.

Consider briefly the algorithms of our codecs.

2400 bps codec. The recorded speech signal (8 kHz, 16 bit) divided by 180 ms windows is passed through third-order high-pass and low-pass Butterworth filters with cutoff frequencies 100 Hz and 3900 Hz. Filtered window is divided into 8 sub-windows for 22.5 ms. Each sub-window is divided into five frequency bands {0-500, 500-1000, 1000-2000, 2000-3000, 3000-4000} with using of sixth-order Butterworth filters. For the first four ranges an indication tone/noise is calculated. If the first two bands have the tone indication the pitch period is calculated. Also for each band the gain coefficient is calculated. Additionally, for the entire sub-window the jitter coefficient and linear spectral pairs are calculated. These parameters are packed into a bit buffer and transmitted to the remote side. At the remote side on the basis of pitch period, jitter coefficient, gain coefficient and tone/noise indicator the excitation signal is created. After interpolation, this signal is fed to the synthesis filter based on linear prediction coefficients calculated from the linear spectral pairs. The output of this filter is the synthesized speech signal.

4800 bps codec. Processed as well as 2400 bps codec signal is divided into six 30 ms sub-windows. For each sub-window the linear prediction coefficients are calculated and converted into linear spectral pairs. Then, the original signal is passed through the inverse filter, based on the linear prediction coefficients. Obtained signal is divide into two parts. For the first part (180 samples) parameters of adaptive codebook initialization are calculated. By the second part (1260 samples) parameters to generate adaptive and algebraic codebooks are calculated. These parameters are packed into a bit buffer and transmitted to the remote side. On the remote side on the base of parameters for adaptive and algebraic codebooks the excitation signal is synthesized. This signal is passed through a filter based on the linear prediction coefficients. The output of this filter is obtained synthesized speech signal.

9600 bps codec. This codec is generally similar to the 4800 bps codec, but it does not use an algebraic codebook. Instead, amplitudes and positions of pulses which substitution into the excitation signal allow minimize the deviation of the synthesized signal from the source are calculated. The algorithm of sequence of pulses calculation is our 'know-how'. Using this sequence of pulses can significantly increase the quality of synthesized speech.

19200 bps codec. The algorithm of this codec is completely identical to the 9600 bps codec algorithm. More high speed of this codec allows increase the number of optimal pulses and, as result, of synthesized speech quality.

Computational cost of our codecs in MIPS (Million Instructions Per Second) are presented in Table 1. The table shows that the computational costs are low, especially in the decoder. This means that our codecs should be used in the web-conferencing systems where the parallel decoding of multiple flows is required.

In 2009 our codecs were shown as Unified Communications TMC Labs Innovation Award Winners (Dinan, M., 2009).

Table 1

Computational cost of codecs

Codec	Encoder, MIPS	Decoder, MIPS
2400 bps	8	3
4800 bps	6	< 1
9600 bps	8	< 1
19200 bps	10	< 1

Recently, due to the improvement of communication channels quality, some wideband speech codecs for 16 kHz speech signal have been developed. Sampling increasing results in a significant increase in intelligibility and naturalness of speech. The classic representative of this group is the codec SILK, which is used in the Skype system. In 2010 we have developed our own wideband codec for 50 kbps which provides excellent quality of the reconstructed speech with extremely low computational cost (Encoder - 0.5 MIPS, Decoder - 0.2 MIPS). This codec we will use in the next version of SAViiMeeting system.

Problem 3. Work in corporate networks.

All real-time data transfer protocols (standard as RTP and special as media transfer protocol for SAViiMeeting system) are working on the base of UDP protocol. But unfortunately, UDP protocol is always blocked in corporate networks. In these networks only TCP protocol and based on it HTTP/HTTPS are allowed and the data is transferred only through caching HTTP proxy. In this regard, I am faced with a problem: how to organize the real-time data transfer with using of a protocol that is not designed for this? I have proposed the following solution.

To transfer each type of data its own TCP-connection is established. Data transmission is organized with using of POST and GET methods of HTTP protocol. In case of bad communication channel using the TCP protocol rapidly increases the data delivery delay and leads to loss of real time. To eliminate this effect, new TCP-connection is created with a specific time interval (in the SAViiMeeting system – 10 sec.). This new connection is used to continue the data transfer. Data that have not been transferred through old connection is removed.

This method can be used in the presence of corporate HTTP-proxy. In this case, no-caching option must be indicated in the request and response headers. HTTPS protocol also can be used.

Exploitation of the SAViiMeeting system is showed an efficiency of the proposed method. The actual link quality, of course, seriously degrades with the deterioration of channel quality, but the external channels in the corporate sector have good quality.

Problem 4. Data transfer through poor communication channels.

In contrast to the corporate sector, in private sector there are practically no problems with UDP protocol blocking, but a problem of a channel quality is often present. Of course, if the channel quality is completely very bad, the real-time communication is impossible. However, users with such channels should be able to download and view archives.

In contrast to the 'live' conference, playback does not require the transmission of data in real time and it can be implemented using the TCP protocol. However, the data transfer speed with using of this protocol seriously degrades with deterioration of the communication channel quality. It is caused by the fundamental flaw of the TCP protocol, when in the delivery confirmation packet sent only the expected number of the next packet. So, if for example, the packet number five was lost and six, seven and eight were not lost, all of them will be re-transmitted, because the transmitter does not have information about their successful delivery.

To eliminate this problem and reduce useless data transfer through poor communication channel, I have proposed a proprietary protocol based on the UDP which provides a guaranteed data delivery. The essence of it is that for each received packet a confirmation with information on eight previous packets is sent. If the packet is delivered, it sets a flag 1, if not - 0. Transmitter analyzes this information and resends only undelivered packets.

Testing of this protocol is showed that, in comparison with the TCP protocol, a real data delivery speed provided by them is decreased significantly slowly with deterioration of channel quality.

Conclusion

In this paper I described the structure of SAViiMeeting system and give a brief description how to use this system. The system contains server part which consists from the following services:

- web-portal;
- database;

- database-proxy;
- media-server;
- http-proxy;
- sip-gate;
- mobile-proxy.

The client part is presented by clients for Microsoft Windows, Microsoft Windows Phone, BlackBerry and Android. Also some technical solutions which have been developed for this system, for example:

1. method of heterogeneous data synchronization;
2. frame independent audio codecs;
3. method of real-time data transfer through HTTP;
4. UDP-based reliable data transfer protocol.

SAViiMeeting's audio codecs have been noted in 2009 Unified Communications TMC Labs Innovation Award Winners (Dinan, M., 2009) and SAViiMeeting system in whole has been named as Unified Communications Product of the Year in 2008 (Viscusi, 2009) and 2009 (Harrison, 2010). The system is available for to thirty days of trial use and exploitation. The link is http://www.saviimeeting.com.

References

Dinan, M., 2009. 2009 Unified Communications TMC Labs Innovation Award Winners Announced. Available at: http://news.tmcnet.com/news/2009/05/21/4191731.htm, 20.03.2013.

Harrison, E., 2010. TMC's Unified Communications Magazine Announces 2009 Product of the Year Award Winners. Available at: http://unified-communications.tmcnet.com/topics/unified-communications/articles/81558-tmcs-unified-communications-magazine-announces-2009-product-the.htm, 20.03.2013.

Machovikov, A. and Stolyarov, K., 2001. System for Speech Communication through Internet. In: Matousek V., Mautner P., Moucek R. and Tauser K. (eds) Text, Speech and Dialogue. *The 4th International Conference, TSD 2001*, Zelezna Ruda, Czech Republic, Proceedings. Lecture Notes in Computer Science 2166 Springer 2001, ISBN 3-540-42557-8, pp. 229-236.

Viscusi, S., 2009. TMC's Unified Communications Magazine Announces 2008 Product of the Year Award Winners. Available at: http://www.tmcnet.com/news/2009/03/12/4051798.htm, 20.03.2013.

Systems analysis of webpage integration frameworks into mobile devices' applications

Nikolajs Bumanis

Latvia University of Agriculture, Faculty of Informational Technology , Liela Str. 2, Jelgava, LV-3001, Latvia
my.way@inbox.lv

Abstract: *Web technologies allow developing fully functional mobile devices' application with basic knowledge in web development technologies, such as HTML, CSS and JavaScript. The evolution trend of the web technologies in mobile application development led to accepting the cross-platform development approach as the main goal for almost any nowadays projects. Nowadays, developers are able to choose between multiple web technology based frameworks and libraries not only for building the application itself, but also for ensure full customization of application's content, UI styling and functionality. After completing the research was determined that every other real-life project will require a bit different approach at choosing the correct technologies, and it is not purely one-sided decision.*

Keywords: PhoneGap, RhoMobile, jQuery Mobile, Sencha Touch, cross-platform, mobile development frameworks.

Introduction

Nowadays while developing the mobile devices' applications (further in text – application) developers stand before choice problems, associated with selection of web technologies and combinations of them, as well as with selection of the development path itself. Existing web technologies and solutions based on them, such as frameworks and cloud builders, allow developing very specific mobile devices' applications, as well as mass market oriented applications. Often, every other project is able to use new web technologies to fulfil the needs of project, based on applications markets' objective audience, exploitation objectives and business collaboration elements. Business collaboration possibility and constraints can hardly influence the development process, as well as selection of development technologies.

One of the main aims of application development is to ensure applications exploitation availability for m-commerce (Alqahtani, 2012). Thanks to m-commerce product and service distributors can expand their activities, which are limited by client satisfaction factor, by providing new products and services to their clients, based on client deal history. It is commonly accepted, that client gets to know about new product or service from distributor, but the product or service itself could be externally provided and integrated from provider side. By business means exists a problem, when provider focusing on one individual service or product delivery or development often has relatively small client base, which leads to applications development being inconvenient, because expenses on development itself would be much higher than final products potential exploitation profit. Distributor, using their client base can supply providers' products and services to their clients, where chosen providers' products and services are strictly associated with client deal history. Formally, this type of business collaboration is highly attractive for providers, because of possibility to expand on the market, by physically just paying compensation for client base usage.

For m-commerce purpose and for collaboration between distributors and providers application must be developed, so that it could fully satisfy developers' and providers' needs by not only functionality requirements, but development and collaboration expenses' means as well. Application design must live up to client requirements from such view points as attractiveness, usage simplicity. Applications' security and accessibility are one of the main factors as well (Alqahtani, 2012).

Applications can be developed using existing development frameworks, or simply web technologies, which allow getting required results. As popular web development frameworks can be stated: xUI, PhoneGap, RhoMobile and ApplicationCraft (Corral, 2012b; Firtman, 2012). Frameworks are based on commonly used web technologies, such as HTML5, JavaScript and CSS. For applications user interface styling are used frameworks or libraries, such as jTouch, Jo, jQuery Mobile and Sencha Touch (Firtman, 2012).

The aim of the research is to decide, which web technologies or frameworks are more suitable for application development, where applications consist of multiple web pages integrated together, and every web page is developed by distributor or provider, to meet the needs of both sides. In the framework of paper, cross-platform development approach, two popular frameworks – PhoneGap and RhoMobile and two user interface visualization and configuration javascript libraries – jQuery Mobile and Sencha Touch will be analysed, and finally, the criteria for webpage integration technologies selection will be created.

Materials and methods

Concept of crossplatform frameworks

For application development three main development paths can be selected, each defines the final products' functionality, visual styling, security and accessibility. Application can be developed in three ways – as native mobile application, as web application or as hybrid application. Each of these ways has its own strong and weak points. Native application is the best in sense of functionality, precision and overall user experience, but development of such application needs a lot more time and skill, as well as mobile platform specific development skills. Web applications are very fast and easily developed, but have highly constrained functionality and security possibilities. Third option – hybrid application (Fig. 1), are objected to use strong features of two first ways, by combining native development containers with fast and easy web coding (Beckman, 2012).

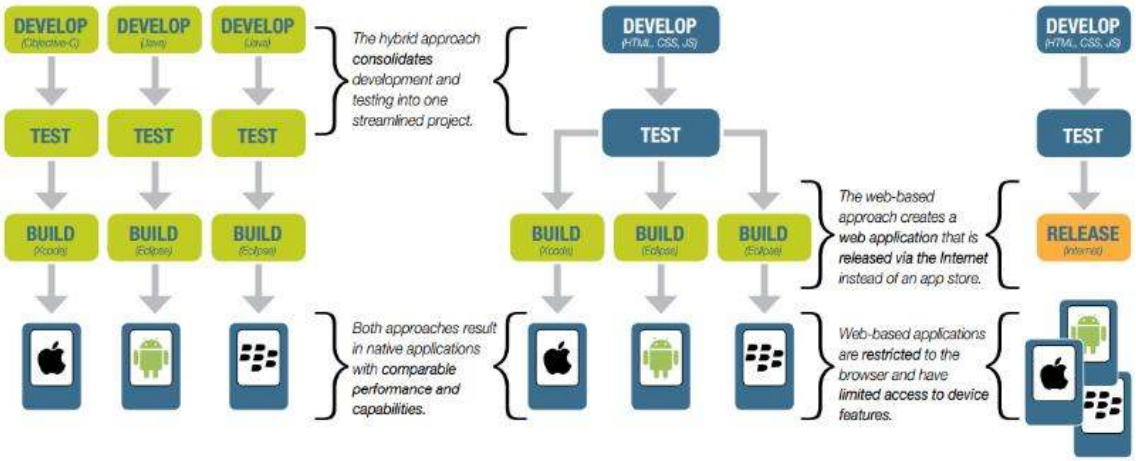


Fig. 1. **The continuum of mobile application development** (Christ, 2011)**.**

While developing application developers must take in mind mobile device's limitations (resources, input possibilities, screen features and etc.), must realize marketing requirements for applications' deployment and modelling, as well as distributors, providers, technologies providers and end user relations (Corral, 2012a).
Nowadays platform dependent applications are common, almost all Apple applications are natives. Developing application for specific platform allows developers to use full functional potential of mobile device, best user-system interaction, while lowers the possible user count because of clients, who use different platforms. It is possible to increase user count by developing the same application for another platform, or by porting the native application using objective platforms' API. Increment of user count using this method requires a lot of time and financial investments. Typical development process (Fig. 2.a) requires native code translation into every other platform's understandable programming language and low level deployment modification, so mobile device can recognize application as native. When coding is finished, application is rebuilt to accept required execution procedure way (Corral, 2012a).

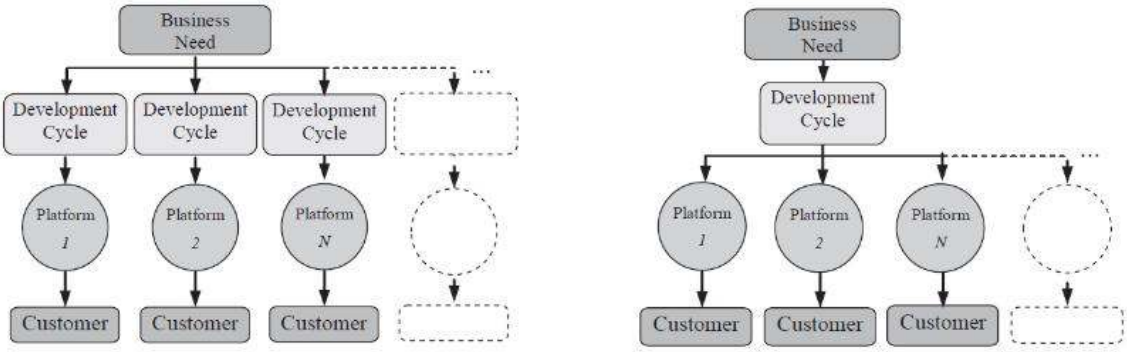


Fig. 2. (a) **Traditional development path**; (b) **Cross-platform development path** (Corral, 2012a).

Thanks to the latest development evolution trends development frameworks were created and they allow adapting existing application to, essentially, almost every popular mobile platforms (Mobile frameworks, n.d.) by going through only one development cycle (Fig. 2.b.). Essentially, using cross-platform development path developers are going through only one development cycle and result of that is deployed on required platforms by choosing objective specified adaptation (Corral, 2012a).

Idea of developing application only once and then deploy on every required platforms has great potential (Duarte, 2011). The only downside is that applications created with the use of frameworks are working on middle layer, what mobile platform is not used to do, thus applications' performance is relatively low compared to native application (Duarte, 2011). Despite of web technologies frameworks' drawbacks, using those it is possible to develop functionally better application then web application, because frameworks allow using mobile devices' features.

PhoneGap framework

PhoneGap framework is one of the hybrid application development tools, which uses before created HTML5 packages and expands HTML5 features using APIs, which grant access to mobile device's specific features, such as camera, geolocation and compass.

PhoneGap is free open code mobile framework which supports six most popular used mobile platforms - iOS, Android, BlackBerry, Windows Phone, Symbian and Bada (PhoneGap supported features, n.d.). PhoneGap allows to authorize native mobile application by using well known programming languages – HTML, CSS and JavaScript, allows connection to mobile device's specific features using platform independent JavaScript API and rebuilds application in the way, so it can be deployed and distributed using multiple different web stores and markets (Dingson, 2011). PhoneGap is package of libraries with the main purpose to "build bridge" between native mobile platform API and JavaScript "hook" – the same for all mobile platforms API is webpage's view – some way of representing webpage on the mobile device's screen(Building mobile apps, 2011). PhoneGap's approach is to use mobile device's web browser as an abstract intermediary layer, which allows implementation on JavaScript based logic layer and presentation layer, built on HTML and CSS. Same as desktop processing, this structure is easy to transfer on different web browsers; however this allows creating only script based applications, working web browser runtime, and thus, JavaScript cannot fully use mobile device's features (Corral, 2012b).

To solve this problem PhoneGap offers API package, which allows managing lower level components, such as telephone handling and hardware features, by using native engine (Corral, 2012b). API becomes available to JavaScript after it is loaded into web browser through PhoneGap's JavaScript engine. Thus, developers must rely on using web technologies (HTML5, CSS, JS) and web programming, because logic layer will count every excessive extension and interface, which influence the overall application resource consumption and mobile device's performance (Fig. 3).

Fig. 3. **PhoneGap application architecture** (Corral, 2012b).

PhoneGap advantages:

- Free (whole framework is open source project);
- Ability to deploy a single codebase on multiple platforms;
- Interfaces with mobile device's hardware such as camera and GPS by using JavaScript;
- Integrated with native first code (application based web browser's object);
- Support all major mobile platforms;
- Applications are built using common web technologies such as HTML and CSS;
- Cloud based compiler;
- Purchased by Adobe (great support and large community).

PhoneGap disadvantages:

- Application is rendered using platform's web browser engine, not individual native UI objects;
- Requires plug-ins for some mobile device's hardware functionality and data security (Keychain plug-in);
- Each mobile platform requires installing appropriate SDK
- Different installation process depending on mobile device's platform (iOS and Androis, especially (Christ, 2011));
- Purchased by Adobe (Adobe doesn't have any open-source products, however PhoneGap code is published under GPL licence).

Best suitable for:

- Products, which require higher customization options' and development's control;
- Products, which require high number of supported mobile devices' platforms.

RhoMobile Rhodes framework

Rhodes is locally executed, device-optimized mobile applications' framework. Applications developed with Rhodes aim to collaborate with the support of enterprise deal application and is developed (Fig. 4) with the ability to synchronize data in local database, for example, SQLite or HSQLDB, or using overall synchronization support framework – RhoConnect, which is another RhoMobile product with partially free open-source. Rhodes supports such mobile platforms as BlackBerry, iOS, Android and Windows Phone (Rhomobile doc.., 2013; Mobile frameworks.., 2012). Rhodes uses Ruby programming language as base for its projects; combined with MVC (Model – View – Controller) pattern it allows creating very well structured projects. Rhodes also has imbued code generation tool, which allows creating applications without deep knowledge of Ruby language (however, high level projects still need a team of professional Ruby programmers). Same as PhoneGap, excluding Ruby, Rhodes is based on HTML, CSS and JavaScript.

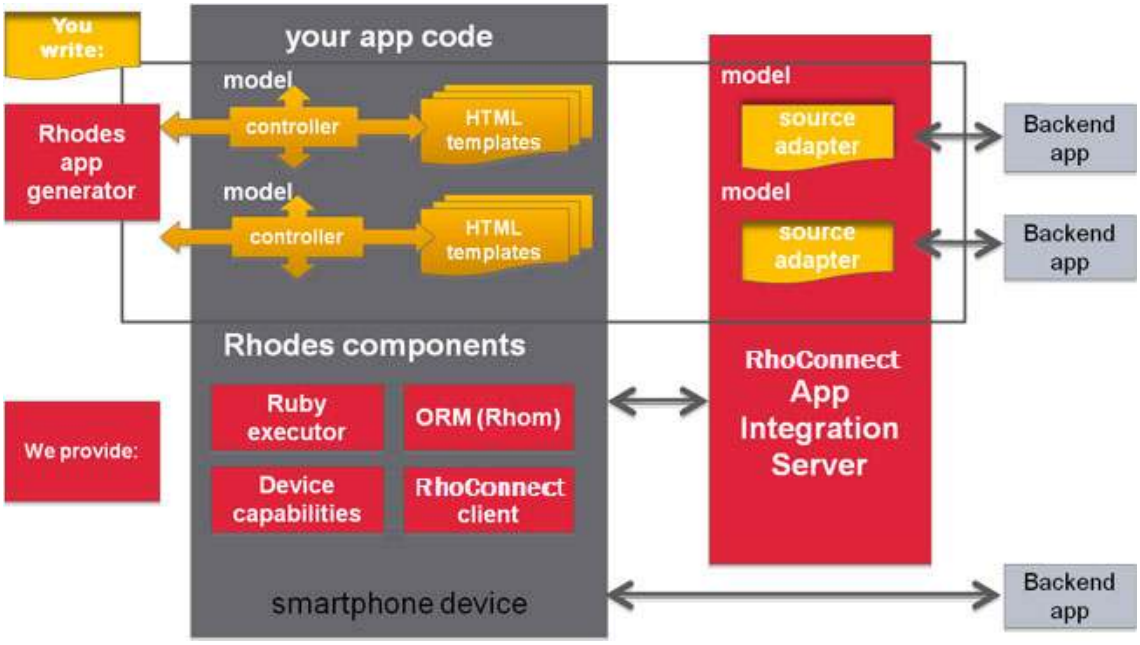


Fig. 4. **Rhomobile Development Environment** (Purdy, 2011.).

RhoConnect is the first framework kind of so called "mobile application integration" server category. Using RhoConnect developers can highly increase quality of compatibility with enterprise deal supporting applications (Rhomobile documentation, 2013). RhoConnect server and RhoConnect built-in client makes all the work associated with data transfer to the device. It is assumed (Rhomobile docucemntation, 2013) that this strategy lowers development efforts for 50 to 80 per cents, by integrating supporting application.

RhoConnect is mainly responsible for going on synchronization (Rhomobile dococumentation, 2013; Christ, 2011; Purdy, 2011): it saves all the data of supporting application into local device, and when supporting application' data is modified it is synchronized with all devices and is available even when device is offline.

RhoMobile Rhodos advantages:

- Partially free (open-source under commerce licence);
- Ability to deploy single codebase on multiple platforms;
- Compiles native application, no web application or hybrid or HTML5 in a mobile browser;
- Interfaces with mobile device's hardware such as camera and GPS by using JavaScript;
- Easy multi-device and cloud synchronizing using RhoConnect;

- Built on Ruby (strength for Ruby programmers);
- Does not require any external SDK installations (Prudy, 2011);
- The only framework with its own security solutions (Prudy, 2011): data encryption, secure authentication;
- Uses MVC model.

RhoMobile Rhodes disadvantages:

- Outputs only native package, not native source code;
- Executes application through RubyVM interpretation package;
- Build on Ruby (weakness for other than Ruby programmers);
- Additional RhoMobile products requires monthly fee (Christ, 2011);

Best suitable for:

- Products, which requires excellent structured project;
- Products, which requires advanced security and data synchronization.

User Interface styling

PhoneGap and RhoMobile very well work with such user interface markup technologies or libraries as Sencha Touch and jQuery Mobile.

Sencha Touch 2 – Upgraded version of Sencha Touch, which was the truly first HTML5 mobile framework created with the aim to fully exploit HTML5, CSS3 and JavaScript abilities to get most flexibility, performance and optimization level (Sencha Touch, 2013).

Sencha Touch (further in text – ST) is developed on JavaScript language, thus JavaScript knowledge is essential to use this framework. ST is the first HTML5 framework created with the aim to expand UI possibilities and compile application for multiple mobile devices' platforms. ST uses MVC (Model – View- Controller) pattern, which grants excellent projects' structuration and overview. Using ST is possible to create fully functional applications for optimized dynamic content and data. ST uses library structure thus is very dependent on global variables. This framework is not the greatest if developers need to provide large amount of themes for their application, because ST really focuses on not only JavaScript code, but also on advanced CSS language – SCSS ("Sassy CSS"), which adds another pack of variables, thus making project very complicated. Mostly, ST is aimed for web kit type browser support, which can be very beneficial with PhoneGap, which uses web kit. The downside on Sencha Touch is that it does not support all major mobile platforms (Mobile frameworks, n.d.), and requires OEM commercial licence purchase.

jQuery Mobile – Based on jQuery and jQuery UI developed mobile framework (physically, just JavaScript library), with the aim to reduce resource consumptions by lowering patch note size and easy-to-build low weight code, progressive user interface enhancement and usability skill wise (jQuery Mobile, 2013).

jQuery Mobile provides great theme creating possibilities by applying on-air build theme using free online ThemeRoller tool. jQuery Mobile supports all major mobile platforms, and essentially, is platform independent JavaScript library (Mobile frameworks, n.d.; jQuery Mobile, 2013). The downside of jQuery Mobile is lower performance compared to ST, and jQuery Mobile is not using MVC pattern, thus can make it harder to structure the project.

Results and discussion

For the analysis purposes following criteria were defined:

- Source code accessibility – Open Source Code, Closed Source Code, Partially Open Source Code;
- Supported platforms – iOS, Android, Blackberry, Windows Mobile, Windows Phone, Symbian, Bada;
- Supported development paths – Native app, Web app, Hybrid app, App generator;
- Supported commonly used web technologies – HTML5, JavaScript, CSS, PHP;
- Use of Model-View-Controller (assuming all technologies support MVC, but only few are based around it);
- Required skill set – HTML, CSS, JavaScript, Ruby, java, etc.;
- Required learning time – low (used only common web technologies), medium (used common web technologies with the help of advanced programming language), high (based around specific programming language, such as Ruby);
- Data security ;
- UI creation and modification – basic (buttons, textboxes, lists), advanced (animation, interactivity, advanced navigation solutions);
- Cloud computing;
- Technical support;
- Technology evolution trend – low activity project (doesn't have recent updates, low ticket response time, bugs are not being repaired), high activity project (have recent updates, fast (1-2 days) ticket

response time, large active community with plug-in ideas and solutions, count of projects, which use this technology), medium activity project;

- Additional features.

Table 1

Comparison of web frameworks

Criteria:		PhoneGap	RhoMobile
Source Code		Open	Open
Supported platforms	iOS	+	+
	Android	+	+
	Blackberry	+	+
	Windows Mobile	+	+
	Windows Phone	+	+
	Symbian	+	+
	Bada	+	+
Supported development paths		Hybrid	Native
Supported web technologies		HTML, CSS, JS	Ruby, HTML, CSS, JS
Use of MVC		-	+
Required skill set		HTML, CSS, JS	Ruby, HTML, CSS, JS
Required learning time		Medium	High
Data security		Plug-in: Keychain	RhoConnect
UI creation and modification		Basic	Basic
Cloud computing		Compilation	Data Synchronization
Technical Support		5 packages, developer count dependent	Forum, Blog
Technology evolution trend		High activity project	Medium activity project
Additional features		-	Data encryption, secure authentication

As a result of analysing two major web development frameworks and two user interface styling solutions two comparison tables were created (Table 1; Table 2). Even though the same criteria were used for both tables, elements cannot be integrated into one table, because they aim to show different purposes. Frameworks, such as PhoneGap and RhoMobile are the main building technologies, which allow to compile and deploy applications on many mobile platforms, but styling solutions, such as Sencha Touch and jQuery Mobile are used only after main application is already complied, modifying content and layout itself by applying JavaScript based functions and queries.

Web technologies are used, mostly, as a set, not an individual solution, for example, PhoneGap can use full potential of jQuery Mobile or Sencha Touch, and developers will choose ui styling solution only after analysing the main web development technology, set of technologies or framework, which is based, essentially, on set of web technologies. RhoMobile Rhodos framework is aimed more towards high quality, enterprise solutions with advanced security possibilities and better data synchronization, even though many projects, essentially, want to include data synchronization as one of the main criteria, and RhoMobile easily completes this task by implementing RhoConnect framework.

Table 2

Comparison of user interface styling solutions

Criteria:		jQuery Mobile	Sencha Touch
Source Code		Open	Partially Open
Supported platforms	iOS	+	+
	Android	+	+
	Blackberry	+	+
	Windows Mobile	+	-
	Windows Phone	+	-
	Symbian	+	-
	Bada	+	-
Supported development paths		Mobile, Web	Mobile, Web
Supported web technologies		HTML, CSS, JS	JS
Use of MVC		-	+
Required skill set		HTML, CSS, JS, MVC	JS, MVC, SCSS
Required learning time		Low	Medium
Data security		-	-
UI creation and modification		Advanced	Advanced
Cloud computing		CMS	Sencha.io
Technical Support		Forum, Blog, Rich Community	Charged ST support, Free Forum and Documentation
Technology evolution trend		High activity project	High activity project
Additional features		ThemeRoller	Dynamic data support

Table data shows that PhoneGap framework is better suitable with jQuery Mobile, because both of them use platform independent source web technologies, such as HTML5, CSS and JavaScript. Even though, PhoneGap can work with Sencha Touch, and the criteria for choosing it may be programmer skill set or license availability. RhoMobile Rhodos framework is built around Ruby, thus requires addition programming skills. However, Ruby language is not the worst disadvantage of RhoMobile Rhodos – it is licensing, which makes mass market products very expensive. RhoConnect synchronization and data security solutions compensate downside. If RhoMobile Rhodos framework was chosen as the main building technology, then Sencha Touch will suit better with it, because of MVC pattern usage, which will grant perfect project's structure and overall overview.

Conclusion

Analysis of only four web technologies and solution touches m-commerce partly, but main web technology evolution trends are well visible:

- Web technology frameworks are developed and modelled using common web technologies, such as HTML5, CSS and JavaScript;
- Cross-platform development path is the leading trend in nowadays mobile application projects;
- Project cannot be truly realized using only individual web technology and set of technologies must be used in almost any cases.

After accomplished analysis, following proposals can be stated:

- Before making analysis and selection of any existing web technology developers must understand the developing applications objectives in the aspect of usage, audience and skill package;
- Each project requires different mobile device's functionality, which can be achieved using different technologies, and technology selection must be appropriated with developers' skills, technology preferences and available licences.

Analysed web technology development frameworks and user interface styling solutions can be used to develop fully functional mobile devices' application with user friendly interface and handling. Analysed technology usage requires only basic web technology knowledge and experience from developers.

Acknowledgements

Research has been developed with support of European Social Fund activity "Support to Master studies" project "Support to the implementation of LLU masters". Contract No. 2011/0020/1DP/1.1.2.1.1/11/IPIA/VIAA/011. Research has been developed in the framework of IT centre of competences branch research Nr. 1.19 "Systems model and prototype development of cross-branch mobile applications and smart card web services' management". Research is being developed in the framework of the first project activity, under contract without compensation.

References

Alqahtani, A., Goodwin, R., 2012. E-commerce Smartphone Application. *IJACSA, Vol. 3, No.8m* pp.54-59.

Beckman, M., 2012. Native, Web, or hybrid: How to choose your mobile development path. Available at: http://www.infoworld.com/d/application-development/native-web-or-hybrid-how-choose-your-mobile-development-path-204079?page=0,0, 26.01.2013

Building mobile apps with Sencha Touch and Phonegap, 2011. Available: http://www.lonelycode.com/2011/05/02/building-mobile-apps-with-sencha-touch-and-phonegap/, 26.01.2013.

Christ, A. M, 2011. Bridging the Mobile App Gap. *Sigma Noblis 11(1) (2011)*, pp. 27-32.

Corral, L., Janes, A., Remencius, T., 2012a. Potential advantages and disadvantages of multiplatform development frameworks – A vision on mobile environments. *Procedia Computer Science 10(2012)*, pp. 1202-1207.

Corral, L., Silitti, A., Succi, G., 2012b. Mobile multiplatform development: An experiment for performance analysis. *Procedia Computer Science 10 (2012)*, pp. 736-743.

Dingson, A., 2011. Writing a Hybrid Mobile Application with PhoneGap and the Dojo Toolkit.

Duarte, C., Paulo Afonso, A., 2011. Developing once, deploying everywhere: A case study using JIL. *Procedia Computer Science 5 (2011)*, pp. 641–644.

Firtman, M., 2012. jQuery Mobile: Up and Running, Sebastopol: O'Reilly Media, Inc.

jQuery Mobile Official Documentation. Available: http://jquerymobile.com/demos/1.2.0/docs/about/intro.html, 27.01.2013.

Mobile Frameworks Comparison Chart. Available: http://www.markus-falk.com/mobile-frameworks-comparison-chart/, 26.01.2013.

PhoneGap supported features. Available: http://phonegap.com/about/feature/, 26.01.2013.

Purdy, G. J., 2011, Rhomobile: Automating App Development across Multiple Mobile Platforms. Available: http://www.mobiletrax.com/Newsletters/tabid/115/EntryId/99/Rhomobile-Automating-App-Development-across-Multiple-Mobile-Platforms.aspx, 26.01.2013.

Rhomobile documentation. Available: http://docs.rhomobile.com/home, 26.01.2013.

Sencha Touch Official Documentation. Available: http://docs.sencha.com/touch/2-1/#, 26.01.2013.

SOFTWARE DEVELOPMENT

A Methodology for Model-Driven Software Configuration Management Implementation and Support

Arturs Bartusevics

Faculty of Computer Since and Information Technology, RTU, Kaļķu str. 1, Rīga LV-1658, Latvia
arturik16@inbox.lv

Abstract: *The paper presents the methodology how to implement model-driven software configuration process in projects of developing applied computer systems and software. A methodology consists few dependent blocks that together allows decrease manual works of implementation and support configuration management process. The most important blocks of the methodology are problems and actions monitoring systems in project and database of knowledge. Monitoring systems can fix all problems in project. Knowledge database consist many rules. Using forward chaining procedure become possible to generate new rules in knowledge database that helps determine is some problems depends on configuration management area and find solution of this problem. New rules generation process help to make corrections in configuration management model. Configuration management model presents all actions and relations needed for implementation of configuration process. Monitoring systems and knowledge database can change current model according to changes in project. The paper describes also risks and limitations of using this methodology and necessary researches to make this methodology useable in projects.*

Keywords: configuration management, model-driven process, knowledge base, monitoring systems.

Introduction

Recent development trends of information technologies show attempts of transforming processes into reuse-oriented and model-driven ones(Osis and Asnina, 2011). The time when success of an information technology project was determined by a number individuals that were masters in their field who would use certain tools to solve particular problems, has come to an end.

Software configuration management is no exception to this. Vast development is taking place in this field. Software configuration management is a discipline that ensures product identification and control of their versions in the course of an information technology project. Another task completed by configuration management is creation of a product and it reaching the customer's operating environment correctly(Hass and Mette, 2003).

Although configuration management has long been defined by various standards of information technologies and various tools have been created to solve particular configuration management tasks, problems still persist. When software is developed, it can be installed in test mode, however in operation or production mode unexpected errors emerge upon installation. An information system is supported by two developer teams placed in different countries. One of the teams corrects errors and the other team develops new functionality. There are situations when both teams need to make changes in the same source code file. The error must be corrected as soon as possible, yet the new functionality will be put in operation only within a month. It is unclear how the changes should be made and stored. The product to be developed successfully and correctly compiles from the programmer's computer, however unexpected errors are encountered during the compilation in the test environment, which are not there on the programmer's computer. Whereas in attempts to set up the product within the operation or production environment, installation is not at all possible(http://experience.openquality.ru/software-configuration-management).

Listed here are only a few of the problems that reveal shortcomings in the field of configuration management. As mentioned above, despite the vast amount of configuration management tools, problems continue to persist. There can certainly be various problem causes in each separate project, however the author believes that the following major causes are usually present:

a. Lack of time and resources to view and introduce configuration management as a discipline, which is particularly characteristic to smaller projects. Often development is started, yet practical issues that emerge due to poor configuration management are only solved when the issue stands in the way to functioning;

b. Absence of a broadly recognized universal plan or methodology that would allow introducing a full configuration management process that complies with IT standards within an acceptable period of time. Often there are people or a single person in the project or company who is able to solve the particular configuration management problems within the project, yet there is no general methodology or a model prescribing how to do it. As a result either the person introduces configuration management only partially and for the purposes of the particular situation or the person is drawn in only when the problem appears.

c. Configuration management often is not perceived as a targeted process, but as concrete tools that "have to be in the project". When asked about the essence of configuration management, some well-experienced IT specialists state a particular version of the control system, continuous integration server, application processing system or another tool this person uses in the project. Without a joint and targeted process new practical problems continue to emerge, and often new problems appear when the old ones are solved(http://experience.openquality.ru/software-configuration-management). This is because there should initially be a targeted process. The author believes at the start of the process goals must be defined, which will prevent concrete problems, and only then can the tools that reach these aims be installed. For instance, before installing a version control system, one should think of how and in which structure the source code will be stored, who will have access to it and whether several parallel branches will have to be sustained, etc.

d. Another problem lies in the project development, namely, the project develops and changes so rapidly that configuration management can no longer solve all the problems and complete all the tasks. The initial configuration management plan then becomes irrelevant, yet there is no time or human resources to bring it up and sustain it. For instance, a company sustains a certain information system and the everyday work is organized in a way that users e-mail their problems to the support team of programmers. While the number of problems does not exceed 100, everything runs well, but what to do when e-mails with issues reach into thousands, and it is found that there are several problems concerning one and the same issue or cases when two or more problems flagged are in conflict with each other? Rapidly introduce one of the application support systems? If so, then how to transfer all the raised problems from the e-mail and how to detect the conflicting ones? In this case there can be many questions, yet it will not be possible to obtain answers rapidly. This is due to the fact that initially the application processing system was not well planned out and it was not taken into account that the project can develop and number of issues may grow considerably(http://experience.openquality.ru/software-configuration-management).

Given these problems that persist in information technology projects, it can be concluded that a methodology is needed that would allow an effective and targeted introduction of configuration management processes into the project. It is needed to advise on more suitable tools for achieving the goals of the particular process and to support configuration management process in accordance with the current needs of the project throughout its course. In other words, the methodology should implement the development of configuration management process and its adaptation to the changing circumstances of the project.

General Description of the Configuration Management Implementation and Support Methodology

Configuration management as a discipline is regulated by several information technology standards, such as ITIL or CMM/CMMI(Hass, 2003). According to these standards a configuration management plan and model can be compiled, which would be suitable for a particular project at a certain time. Afterwards technologies and tools would have to be chosen in accordance with the plan, which would help solve and automate configuration management tasks. When technologies and tools are chosen, these can be installed and the configuration management process can be regarded as implemented within the project. Is that really so? Before answering this question I will mention a few problems that persist in practice and which the author of this article has encountered:

1. When a new project is started, there is lack of human resources and time to study configuration management standards and to compile a configuration management plan and a model for the particular project. Within the actual situation such activities are rarely given enough resources. Most often definition of requirements is started, along with setting up the project and initial development, yet problems that emerge due to poor configuration management, are solved locally without giving much thought to their causes.

2. Even if there are attempts to initially introduce a joint and targeted process of configuration management, the project develops too quickly and soon the initially defined process does not match the actual project situation. As a result the process is defined, however new problems continue to emerge regarding version control, change management, product compilation and installation, and other problems linked to configuration management.

What follows is that an initially thought out configuration management process plan does not yet mean that we can speak of a full-fledged and targeted process that would help avoid problems and obtain a qualitative product. This is also because configuration management is perceived as a necessity to solve several local problems or a process is introduced that quickly looses its effectiveness due to the vast pace of project development. In this case the project often contains a document that kind of describes configuration management, yet it does not fit the actual situation.

All of the above shows that a method is required that would allow doing the following:

1. By using only the available resources, to sufficiently quickly set up a configuration management process, which would comply with a concrete project at a particular period of time.

2. To ensure process development in accordance with the development of the project. Development must be automized within the scope of possibility. That means that the process must change along with the project, and there should be a continuous link between configuration management and other project processes.

By developing the methodology, which could solve the mentioned tasks, two assumptions had to be defined, upon which the configuration management implementation and support methodology will be based, as described in this article. First, each information technology project contains particular actions. The project takes place while there are particular actions. As an example of these actions we could mention development of the specification, writing of the source code and testing of memory, module compilation, preparation of a project report, etc. Second, as a result of actions, certain problems can emerge, For instance, a module has not been compiled, a mistake has been made when testing memory, a programmer has overwritten the changes made by another programmer when modifying the source code. Therefore each project contains actions and problems that emerge when certain actions are performed. Whereas the aim of configuration management is independent of the type of project and technologies used in the project. The aim of configuration management is to ensure product identifications, version control and changes reaching the end product correctly.

If we imagine that there are actions and problems in each project, then the task of the configuration management process is as follows: to minimize the number of problems that emerge by taking actions to reach configuration management goals.

Given the just formulated configuration management task, the following methodology for implementing and supporting configuration management is offered. In the initial stage of the methodology, monitoring of actions and problems is performed. There are two systems that record all the actions and problems. During the action monitoring only those actions are obtained form the set of actions that are oriented towards a solution of configuration management tasks. When actions related to configuration management are obtained, the link between these actions is defined. In the second step of the methodology actions and links between them are transformed into a model. In the following step the model is supplemented with action attributes, which describe actions from the aspect of the chosen tools and technologies. Parallel to action monitoring and transformation, problem monitoring takes place in the model. Each problem is registered in the storage. Later on it is determined due to which actions this problem has emerged. If this action belongs to a set of configuration management actions, a problem solution is sought. In order to find a solution, a base of configuration management rules is required. This base finds solutions to concrete problems and then introduces these changes to the model. Action and problem monitoring takes place continuously. Changes to the configuration management model are made in two cases: if a new action is defined that concerns configuration management and if a solution to one of the configuration management problems is found in the configuration management base of rules. Fig. 1 shows a schematic representation of the methodology, showing components of the methodology and links between them.

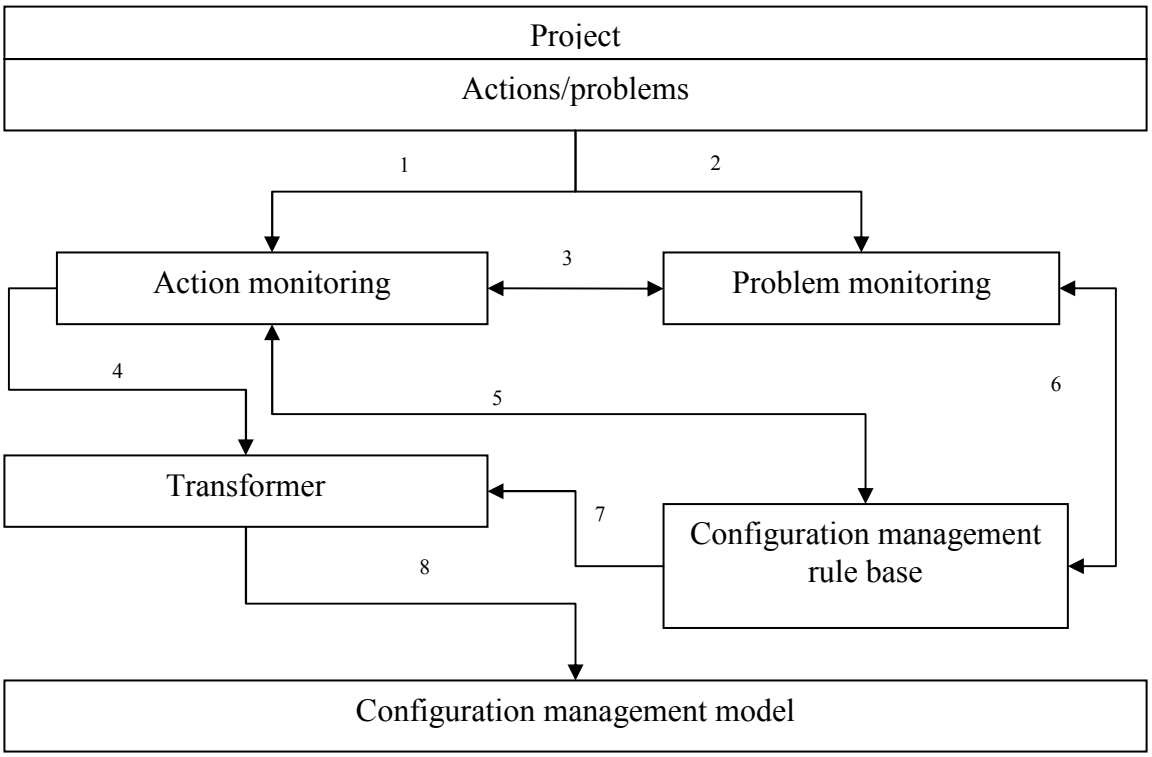


Fig. 1. **Methodology for Developing and Supporting the Configuration Management Model.**

Methodology consists of five elements and eight data flows, which ensure a link between methodology elements. A configuration management model is one of the components of the methodology. The model changes and develops while the methodology is in action. The general goal of the methodology is to sustain the configuration management model in line with the current situation of the project. Whereas the goals of the model is to considerably show to all project participants the configuration management process in a clear way. Another goal of the model is to show in what ways each of the configuration management tasks is solved and to what extent this solution contributes to development of a correct end product, which is one of the most important configuration management goals.

Description of Methodology Elements

Action Monitoring

This system continuously analyzes all actions taking place in the project. The action monitoring system receives information when a certain event has taken place within the project. Afterwards the obtained information is analyzed. As a result of information analysis action is identified. One action in the context of this system is the vector D<I, A, P, S>, which describes one particular action and where:

I – identifier of action,

A – actor that performs this action. This can be a concrete person from a project, for example, a tester, a programmer or a system analyst. The actor in this case can also be one of the systems used by the project, for instance, a continuous integration server, in which the automatic project tests are running,

P – action performed by the actor. This is a concrete verb that clearly characterizes a particular action, for instance, to test, to compile, to write, to register, to generate, etc.

S – a set of action attributes. Attributes help better understand the meaning of a particular action. Attributes contain information about what exactly the actor carries out, which tools it uses, and which systems it works with. At the moment when the action monitoring system contains several actions, the S set can also contain an attribute that shows links with other actions.

When information about the project is processed and the attempt to transform it into a formal definition of action is successful, action-monitoring system sends information about the actions to the configuration management rule base, from which it receives a response whether the action concerns the configuration management process. In case of a positive answer, information about actions is sent further to the transformer – an element that transforms action vector D into an element of the configuration management model. The action monitoring system functions on an ongoing basis.

Problem Monitoring

The principle of problem monitoring system activity is very similar to the action monitoring system. The system continuously analyzes events in the project with an aim to identify errors and problems. When a problem emerges in the project, for instance, changes have disappeared, tests work with errors, source code cannot be compiled, then the problem monitoring system is triggered. The obtained information is analyzed and information about the problem is formalized. Within the given methodology a problem is a vector P<I, A, D, R, R1> , where:

I – identifier of a problem,

A – a noun that characterizes what the problem concerns, e.g., a source code file, a compiler, or a specification. This can also be a person who caused the problem, for example, the programmer.

D – a verb or verbs that characterize the essence of the problem, e.g., did not compile, lost, did not work, etc.

R –reference to an action, as a result of which the mentioned problem emerged. At the moment when the problem emerges, the value of R is empty. It is only filled when a response is obtained from the action monitoring system.

R1 – possible solution of the problem. Initially the value is empty. It is only filled when a response is obtained from the action monitoring system.

When the problem is formalized, information about it is sent to the action processing system. That is where all actions are searched with the aim to identify the action due to which the problem emerged. At this point the attribute of vector R is filled. If the action cannot be identified, the problem is recorded into a special list of the problem monitoring system, which will only be analyzed once the set of actions within the action monitoring system is changed.

After a response is received from the action monitoring system and the problem vector attribute R is filled, the problem is sent to the configuration management rule base, where solution to the problem is sought. If the solution is found, the problem monitoring system receives a response that the problem has been solved and it is marked as resolved. At this moment attribute R1 is filled for the problem. Whereas if no solution is found, the problem is sent from the configuration management rule base with an empty R1. In this case the problem is added to a special list by the monitoring system.

Configuration and Management Rule Base

This element of methodology is of great importance. It determines whether the configuration management model will be able to correctly develop according to the new project requirements and the changing situation.
The configuration management rule base consists of the following elements:

- Action mapping – a map that contains information about actions required to solve one of the configuration management tasks. The map consists of several elements. Each of these elements has two attributes. The first attribute is one of the configuration management tasks, for instance, ensuring version control, yet the second element are the actions necessary for solving this task as well as a list of potential tools and technologies to be used when solving the given task. It is by following this map that allows determining whether action D from the action monitoring system concerns the configuration management discipline or not.
- Problem solution mapping – a map that contains problems and their potential solutions and tools to implement these solutions. This is the map on which search takes place when looking for a solution for a concrete problem P.

The configuration management rule system is hosted by an individual with theoretical and practical knowledge on configuration management. For instance, it can be a configuration manager in the project. The question in what ways to implement the ability to process the mentioned map elements with the help of software remains yet to be answered. It needs to be acknowledged that regardless of map realization, the development and hosting of a qualitative system will not be possible without the intervention of a specialist.

Transformer

This is a tool configuration model building. It receives actions D with attributes and transforms the acquired vector into the model. Realization of the transformer depends on what type of model is used. In any case the model has certain elements and links between them. The transformer receives one action D<I, A, P, S> and it has to be an algorithm that transforms this vector into a model element. Realization of the transformer will not be reviewed in this paper. In order to realize the transformer, additional research and experiments are required.

Configuration Management Model

This model is the main component of the methodology. The general goals of the model are as follows:

- To demonstrate the solution of configuration management tasks as a targeted process;
- To demonstrate what tools and technical means solve which of the problems and in what ways the tools interexchange information.
- To help the configuration manager avoid a situation, in which the solution of one problem is the cause of another;
- To help the configuration manager avoid the emergence of similar problems once a certain problem is solved.

In the context of the described methodology any project problem that refers to one of the configuration management tasks is a shortcoming or an error within the configuration management process model. Therefore by fixing this shortcoming or this mistake, first of all the particular problem is solved and second of all theoretically a situation is achieved, in which a similar problem will not repeat itself. It is therefore crucial for the model to be as precise as possible, while the process itself should fit the particular project as a much as possible.
This paper will offer a configuration management process model, which is based on actions from the action monitoring system. Upon receiving the action vector D, the transformer develops the following model components:

1. Action – the main element of the model, which reflects a particular action. This element reflects attribute P from the vector D.
2. Object (actor) – an element that reflects attribute A of vector D. It is a person or an object that performs action P.
3. Action attributes – elements that reflect attribute S of vector D. These are attributes characterizing action, which clarify the meaning of action, show in what ways the tools are used when performing this action, as well as what kinds of interfaces exist for communication with other tools (attributes).
4. Report – an element that contains information or a set of transfers that emerge after actions are completed. It is a report that is sent to another object (actor), which belongs to a different action.

Therefore the model of the configuration management process consists of four basic elements. The main element is action. Action can have one or several points of entry, in which the object (actor) activates this action by giving all necessary information. Action has a characterizing set of attributes, which gives full understanding on the course of action and its tools, helping to complete or to automize this action. Action also has one or several points of exit, which contain the report that, in itself, can either be a transfer of a particular action or a set of

information for another object (actor). This information may be necessary for another object (actor) in order to activate the next action of the model. Fig. 2 shows a fragment of the model in the context of a single action.

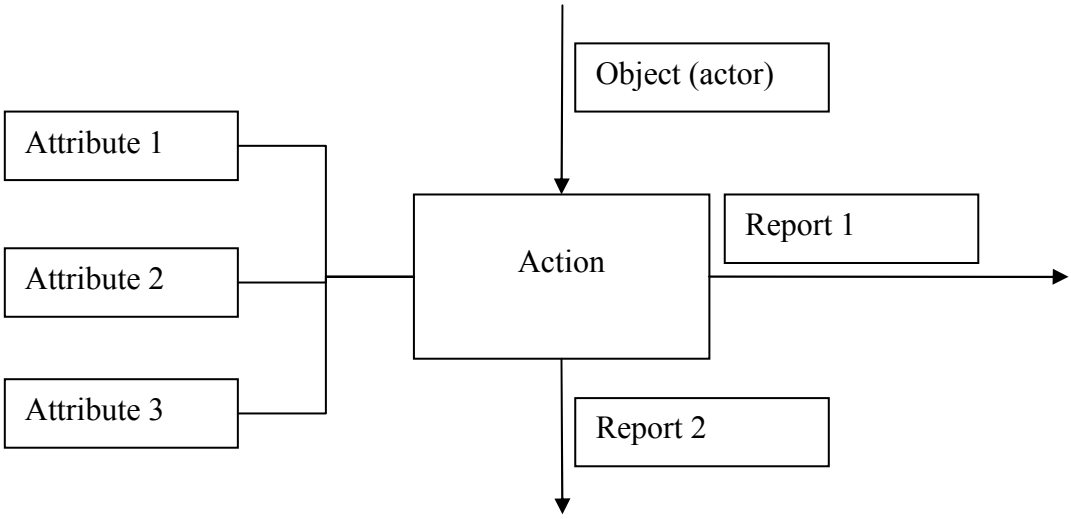


Fig. 2. **Fragment of a Configuration Management Model.**

Tools and Obstacles to Introducing Methodology

When discussing the possibility of making use of the described methodology in current practice, the author detects several risks and obstacles to implementing the practical approbation of the methodology. There is potential of the following risks:

- Monitoring systems can be imprecise or may not be suitable for the particular project. Therefore the model development process in line with the methodology may be imprecise or contain errors. If monitoring cannot identify certain actions or problems significant from the aspect of configuration management, it will not be possible to develop a precise model corresponding to the real-life situation.
- The configuration management rule base may have shortcomings and it may be imprecise, which may lead to inability to classify the actions registered in the project. As a result the model may contain a certain action that does not at all correspond to configuration management and rather is related to a completely different project discipline. The configuration management rule base may also create obstacles to problem solution, for it may not contain a solution for a particular problem. As a result the problem will remain to exist within the project and the model within a model will not detect a possible solution. The configuration management rule base may also contain a faulty fact, which is based on a subjective assessment of a particular specialist, yet which in reality is untrue.
- There is a risk that the formalization of the problem solution will not be of sufficient quality in order to correctly introduce changes to the configuration management model. The idea is that it is not enough to find a solution to the problem, for one must also present this solution to the model transformer in a clear and unambiguous way in order to make correct changes to the model. If a solution is found in the rule base, but in the course of solution formalization mistakes are made, the changes in the model may not solve the problem and it may repeat itself in the future.

Another risk group has not just to do with methodology, but with the current situation within the discipline of configuration management. IT specialists, instead of perceiving configuration management as a targeted process that solves tasks of a certain kind, may sometimes perceive it as a total of tools and technologies, without which the “project will not work”, according to them. Such specialists are not likely to try something new and view configuration management as process- and model-oriented. The author believes that sometimes due to psychological reasons it is difficult to choose something completely new and unknown to this point, for in the precious years of the existence of the IT field there has been a tendency to perceive configuration management in the context of tools. Yet taking into account the modern tendency for the field of information technologies to establish itself as a science and processes to become model-driven, to perceive configuration management solely as a set of several tools is incorrect and not in line with the current tendencies(Frakes, 2005).

Conclusion

Given that the methodology is new and there is no detailed description of its elements and implementation, the author believes that we still have a long way to go to achieve practical approbation of methodology and result assessment. Therefore several tasks must be completed, which should not just strengthen the methodology, but also create possibilities for applying it in practice. Given the current situation, the following future tasks are at place:

- To work on possibilities of implementing and introducing a monitoring system. It is necessary to study the existing monitoring systems and to draw conclusions which of them would be the most suitable for action and problem monitoring. One problem to be solved is signal generation, and the other problem – what information must be in the signal in order to formalize it into an action or problem vector.
- Working with the configuration management rule base. As mentioned before, the rule base is an essential component of the described methodology. For the rule base to be able to solve various problems, it should contain many rules, facts and a summary of best practice examples. Given the fact that configuration management is currently being described by many sources of information and is implemented in practice by numerous tools, development of such a rule base manually would be a lengthy and irrational process. As noted before, one of the advantages of the described methodology is the ability to apply it from the very start of the project, without investing additional resources into preparatory events. And if an extensive configuration management rule base is to be created manually, the methodology looses its meaning. From here we can conclude that we need a sufficiently fast and automatic way to create a configuration management rule base. We therefore need to search for ways how to achieve this. The author allows the possibility of having to use one of the data acquisition or artificial intellect methods, or possibly even several methods at once. Some thought should be given to depiction of rules in order for the monitoring systems to work with the rule base.
- Given the existing tendency to make the IT processes reuse-oriented and model-driven, it can be concluded that the attempt to make the configuration management process into a model-driven one is modern and corresponds with the tendency of the field of information technology to reinforce itself as a scientific field.

References

Hass, J. and Mette, A., 2003. *Configuration Management Principles and Practice,* London: Addison-Westley.

Frakes, W.B. and Kang, K.C., 2005. Software Reuse Research: Status and Future. In: IEEE Transactions on Software Engineering, vol. 31., no. 07., pp. 529-536.

Osis, J. and Asnina, E., 2011. Model-*Driven Domain Analysis and Software Development: Architectures and Functions,* IGI Global, Hershey - New York, 518 p.

Yuri Udovichenko, 2011. Change management. Available at: http://experience.openquality.ru/software-configuration-management, 13.02.2011

Development of BRAINTOOL for generation of UML diagrams from the two-hemisphere model based on the Two-Hemisphere Model transformation itself

Oksana Nikiforova, Olegs Gorbiks, Konstantins Gusarovs, Dace Ahilcenoka, Andrejs Bajovs, Ludmila Kozacenko, Nadezda Skindere, Dainis Ungurs

Riga Technical University, Faculty of Computer Science and Information Technology, Meza street 1/3, Riga, LV-1048, Latvia
oksana.nikiforova@rtu.lv; olegs.gorbiks@rtu.lv; konstantins.gusarovs@rtu.lv; dace.ahilcenoka@rtu.lv; andrejs.bajovs@rtu.lv; ludmila.kozacenko@rtu.lv; nadezda.skindere@rtu.lv; dainis.ungurs.@rtu.lv

Abstract: *The Unified Modeling Language (UML) is an industrial standard for object-oriented software specification, which offers notational conventions for system modeling at the initial stage of software development. An actual problem is to develop a tool available for automatic generation of UML diagram from some form of the problem domain description. Authors have been introduced such a tool, called BrainTool and developed by a research group in Riga Technical University. BrainTool supports transformation of the problem domain presented as two-hemisphere model into several kinds of UML diagrams and give an ability to export these diagrams for further software development into UML compatible modeling tools and IDEs. Present paper demonstrates the comparison results of the manual UML class diagram creation during the development of the tool, automatic generation of the UML class diagram received by application of BrainTool for the BraintTool itself and so called "to be" class diagram for BrainTool development. These results give a basis to discuss about possible improvements of the transformations offered by two-hemisphere model and potentially "richer" usage of the model for generation of the UML diagrams suitable for further software development.*

Keywords: model transformation tool, two-hemisphere model, UML class diagram, BrainTool.

Introduction

The object-oriented approach is widely used in software development. One of the tasks of software development is to present different aspects of the system for the implementation of the software solution for the required system. In solving this task, system modeling became an important activity in software development. The goal of system modeling is to represent the system graphically, in a form understandable to analysts, developers and at least partly understandable to the customer. A systematic approach to the derivation of the system model from information about the problem domain plays an important role in completing the task of system modeling. Moreover, the increasing interest towards software development within the framework of Model Driven Development (MDD) (Stahl and Volter, 2006) turns the focus again to the area of model transformation at different levels of abstraction. Unified Modeling Language (UML) (OMG's UML) is an industry standard for software specification and modeling in an object-oriented manner. The UML class diagram is used to model class specification and serve as a "bridge" between the information about the problem domain and the information required for definition of the software components and their architecture. Currently, researchers are trying to achieve a high enough level of automation in creation of the UML class diagram and derivation of the diagram from information about the problem domain.

There exist a number of tools which generate the UML class diagram. Some of them enable to define several elements of class structure based on data presentation of the problem domain. Others generate a class diagram from existing source code, to display the structure of the developed system. However, the problem of automatic generation of the UML class diagram from the formal and still customer-friendly presentation of problem domain is not solved yet. Authors have been presented the tool, called BrainTool at the previous AICT conference (Nikiforova et al., 2012a). BrainTool (Website of BrainTool), developed by researchers of the Riga Technical University, is a step forward in the area of automation of the modeling process. Authors of BrainTool propose to generate UML class diagram from the so-called two-hemisphere model (Nikiforova and Kirikova, 2004) of the problem domain, which presents information about processes, information flows between these processes and pre-defined types of these information flows.

Present paper demonstrates the comparison results of the manual UML class diagram creation during the development of the tool, automatic generation of the UML class diagram received by application of BrainTool for the BraintTool itself and so called "to be" class diagram for BrainTool development.

The paper is structured as follows. The next section describes the related work in the area of UML class diagram generation and tools supporting this generation. Section 3 remains the main components of BrainTool and the transformations used for generation of the UML class diagram from the two-hemisphere model. Section 4 gives

comparison results and approves the correspondence of the generated class diagram to the “to-be” solution. Several conclusions and directions for future research are stated in the fifth section.

Related work

Since the beginning of the 1980s a great number of modeling tools and model generating software systems have been offered to attack problems regarding software productivity and quality (Balzer, 1985). Modeling tools developed since that time were oversold on their "complete code-generation capabilities" (Krogstie, 2005). Nowadays, similar situation is observed in modeling tools, using and integrating UML models at different levels of abstraction and automation of software development (Mellor et al., 2004).

Most of today's tools combine a number of modeling and code generation functions in a more or less open fashion. The traditional modeling tools provide a model editor and a model repository. A code generator based on a scripting language and plugged into a modeling tool provides the transformation tool and transformation definition editor. In that case, the transformation repository is simply text files (Kleppe et al., 2003).

The variety of the "model-driven" tools can be divided into tools created for defining the system model itself and tools to support code generation from the UML model. The first group of tools is so-called "UML editors", where the developers of these tools provide different levels of automation of the actual model creation. BrainTool demonstrated in this paper can be classified as a tool for automatic creation of UML class diagrams, where the result of the generation – the UML class diagram – is importable either into UML editors for further refinement and working with model or into code generation tools for further generation of software components.

Loniewski, Insfran and Abrahao (2010) describe the results of a survey about different approaches used for transformation of system requirements to system design and implementation. The survey shows the result of analysis of different approaches to transformation of the problem domain description into the UML class diagram during the last 10 years, published in four digital libraries (IEEEXplore, ACM, Science Direct, Springerlink). The survey states that there exist many approaches with different types of solutions for the generation of a UML class diagram. Moreover, the authors analyze the approach based on several criteria, one of them is tool support. Analysis of the automation level in these approaches shows that 25 out of 71 approaches described in corresponding papers are supported with a tool. However, Loniewski, Insfran and Abrahao (2010) stress that these tools are academic tools and are not widely practically used as far as they are created to approve the automation level of the approach offered by their vendors.

One more kind of the related tools are tools that generate the class diagram from a data structure or a data model. These tools require a solid contribution from a software specialist to define all these structures. It is already the modeling of the UML class diagram itself. In contrast to these tools, BrainTool generates the class diagram from initial information about the system, which is understandable for the business analyst and doesn't require software knowledge for its modeling. Therefore, a tool that generates the class diagram in the initial stages of the project is very useful. It allows to automatically create a static structure of the developed system and serves as a base for further code, avoiding mistakes and mismatches between requirements and implementation.

As far as for the evolution of the two-hemisphere model, which is a base model for generation of the UML class diagram in BrainTool, the main idea of displaying the initial information about the system with two interrelated models– the business process model and the concept model – was introduced by Nikiforova (2002). The title of the approach as the two-hemisphere model driven was defined by Nikiforova and Kirikova (2004), where the hypothesis about how to use two interrelated model to share the responsibilities between object classes was demonstrated on the abstract example and later in a real project (Nikiforova et al., 2006). In both cases the two-hemisphere model was created manually, in the first one by authors and in the second one by an independent problem domain expert. Successfull application of two-hemisphere model transformation into the UML class diagram served as a motivation to support these transformations by software system. The first software prototype of tool supporting two-hemisphere model based transformation was introduced in 2008 (Nikiforova et al., 2008). The prototype used textual information in special format as a source and produced a text file containing description of the resulting UML class diagram as a specification, where classes, attributes, methods and relationships were listed in pre-defined format. Analysis of these generated text files gave authors an ability to refine transformations for definition of relationships between classes, the results are published in (Nikiforova and Pavlova 2009). Currently, the ability to apply the two-hemisphere model for generation of the UML sequence diagram with attention to the timing aspect is investigated and preliminary results are published in (Nikiforova 2010). So far, the continuing research in the area of model-driven software development and an increasing demand in the industry for automation of the ability to bridge the gap between problem domain and software components, can serve as a motivation to develop the first version of BrainTool, which gives an ability to draw the two-hemisphere model in the manner suitable for the problem domain expert and to generate the UML class diagram from it.

Moreover, instead of manual creation of the UML class diagram directly from information about problem domain based on principles of object-oriented analysis, the proposed BrainTool gives an ability to use already existing business artifact – a business process diagram is widely used in many enterprises, and the structure of information flows between processes is definable under description of user stories. A lot of organizations are

using different tools for business process analysis and therefore they have complete and consistent models of their organizational structure, employer responsibilities, business processes and the structure of documentation flows, in other words, well-structured initial business knowledge, which can serve as a basis for even automatic creation of the two-hemisphere model.

The main benefit of the two-hemisphere model is that it can be created and often already is created by the business analyst at the customer's side. A Standish group survey shows that about 83% of companies are engaged in business process improvement and redesign. This implies that many companies are very familiar with business process modeling techniques or at least they employ particular business process description frameworks (Rittgen, 2010), (Peyret and Miers, 2010). On the other hand, the practice of software development shows that functional requirements can be derived from the problem domain description as much as 7 times faster than if trying to elicit them directly from users (Aalst, 2007). Both facts mentioned above and the existence of many commercial and open source business modeling tools are a strong motivation to base software development on the business process model, rather than on any other soft or hard models.

Therefore, with minimal efforts created and intuitively understandable by customer two-hemisphere model can be used for automatically generating class diagram prototype that can be later reviewed and used in software development.

Implementation of two-hemisphere model driven approach in BrainTool

The two-hemisphere model driven approach is based on the transformation of two interrelated models into the UML class diagram. These two initial interrelated models are: business process model (shortly – process model), which displays behavior of the system, and the model of conceptual classes (shortly – concept model), which displays a skeleton of system's static structure. The meaning of objects in an object-oriented philosophy gives a possibility to share responsibilities between objects based on the direct graph transformation, where the data flow outgoing from the internal process in the process model becomes an owner of this process for performing it as an operation in object communication and further is mapped into class responsible for this operation in class diagram.

The current version of the implementation of the two-hemisphere model driven approach can be stated as a standalone tool titled as BrainTool in correspondence with the title of the approach, which in turn is derived from cognitive psychology (Anderson 1995) by analogy with human brain consisting of two interrelated hemispheres. According to Kleppe, Warmer and Bast (2003), a modern trend in system modeling tools is having the components to implement a model editor, a repository, its validation and transformation to another model. BrainTool gives a possibility to create the two-hemisphere model, to save it in the defined repository, to apply all the defined transformations for generation of the UML diagram and to export it in XMI format. Fig.1 demonstrates the general view of BrainTool, transition to XMI file of the generated UNL class diagram, and its import into Sparx Enterprise Architect tool.

Model Editor is a part of the tool providing model creation and modification possibilities. BrainTool is based on a C# implementation using .NET framework 4. Model Repository is the "database" for models, where they are stored. Model Repository is implemented as an XML specification used by both the model editor and the transformation component. An XMI export feature is present, implemented as a Python program. The Transformation Definition Editor is used for transformation definition construction and modification. Currently, the Python interpreter is being used to support this component. However, it is possible to define the transformation in any programming language. The Transformation Definition Repository is storage for transformation definitions, where a set of Python scripts is used for the current implementation. However, the usage of any console based application understanding BrainTool's XML schema is acceptable. And, finally, The Model Validator is a component used to check if the model is well-enough defined and has no potential problems that can affect the transformation result. This component is implemented as a standalone transformation using the Python programming language.

After elements of the two-hemisphere model are transformed into the class diagram, BrainTool gives the possibility to export it in XMI format to be later used in other UML editor or code generators that are able to import UML class diagrams in XMI format. Currently, most UML compatible tools use their own modifications of the XMI format and a developer cannot be sure about the result of import/export (Nikiforova et al., 2011). Therefore, the authors were forced to adjust the exported XMI for the requirements of a specific corresponding tool. The Sparx Enterprise Architect is selected for the experiment, but it is not a problem to define the elements of the UML class diagram according to the specific requirements for import in any other UML tool.

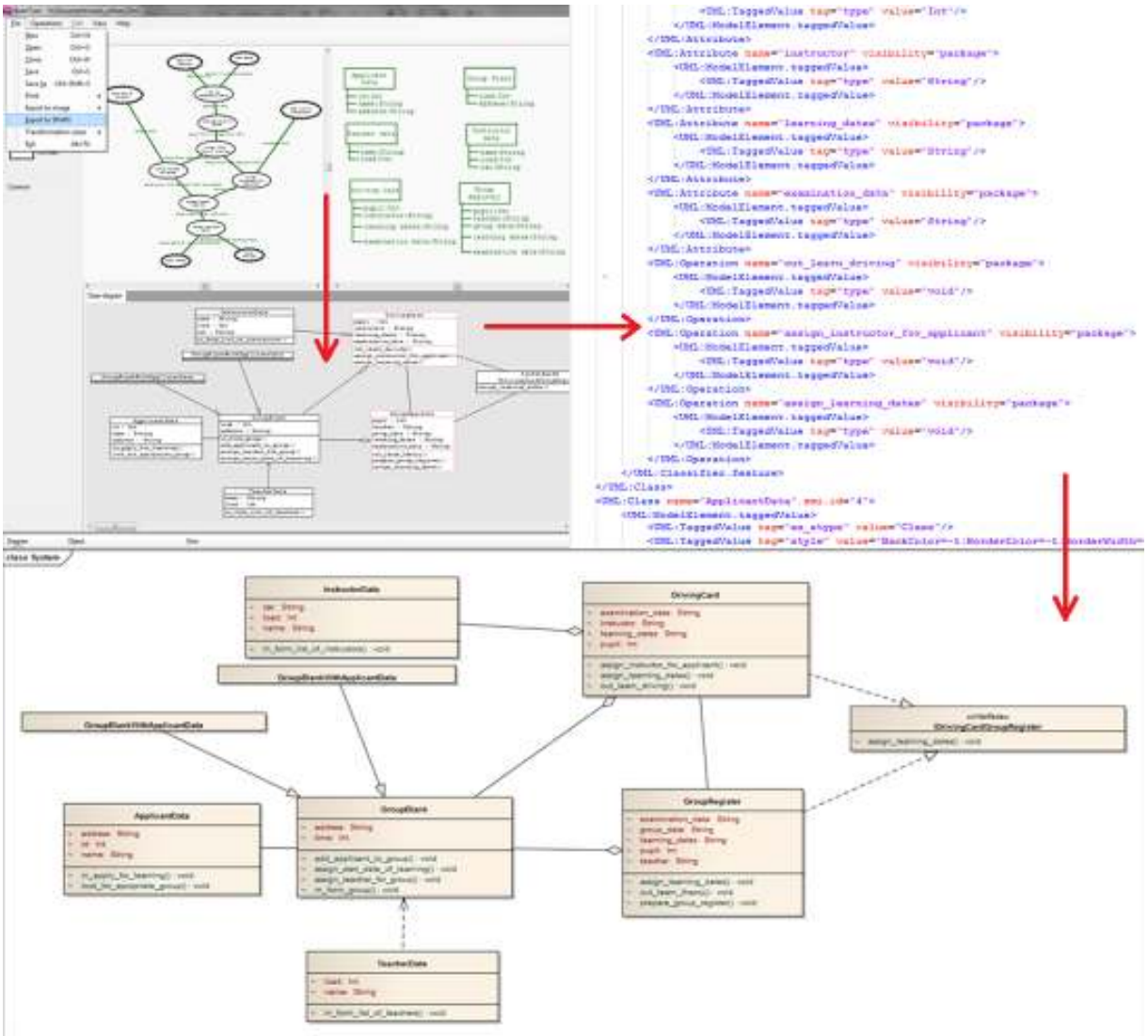


Fig. 1. **General view of BrainTool and import of the generated UML class diagram into the Sparx Enterprise Architect tool.**

Evaluation of the generated UML class diagram

Development of BrainTool gives a possibility to create the two-hemisphere model for different problem domains in turn to analyze possible improvements of the two-hemisphere model driven approach and to approve the correspondence of the generated UML class diagram to the practical usage for software system developments. Authors have been developed the two-hemisphere model, where the problem domain is the two-hemisphere model driven approach itself in order to develop BrainTool. The modeler of the problem domain should in first to model any process of the operating system; then some of the other processes have to be placed to unable the definition of the information flow between any two processes. Information flow, at first, should be created, linked to processes and then it is possible to define the data type for this information flow, etc. Fig. 2 demonstrates a class diagram generated by BrainTool from this two-hemisphere model. Fig.3 demonstrates a class diagram created during development of BrainTool itself and using iterative software development process and best practices.

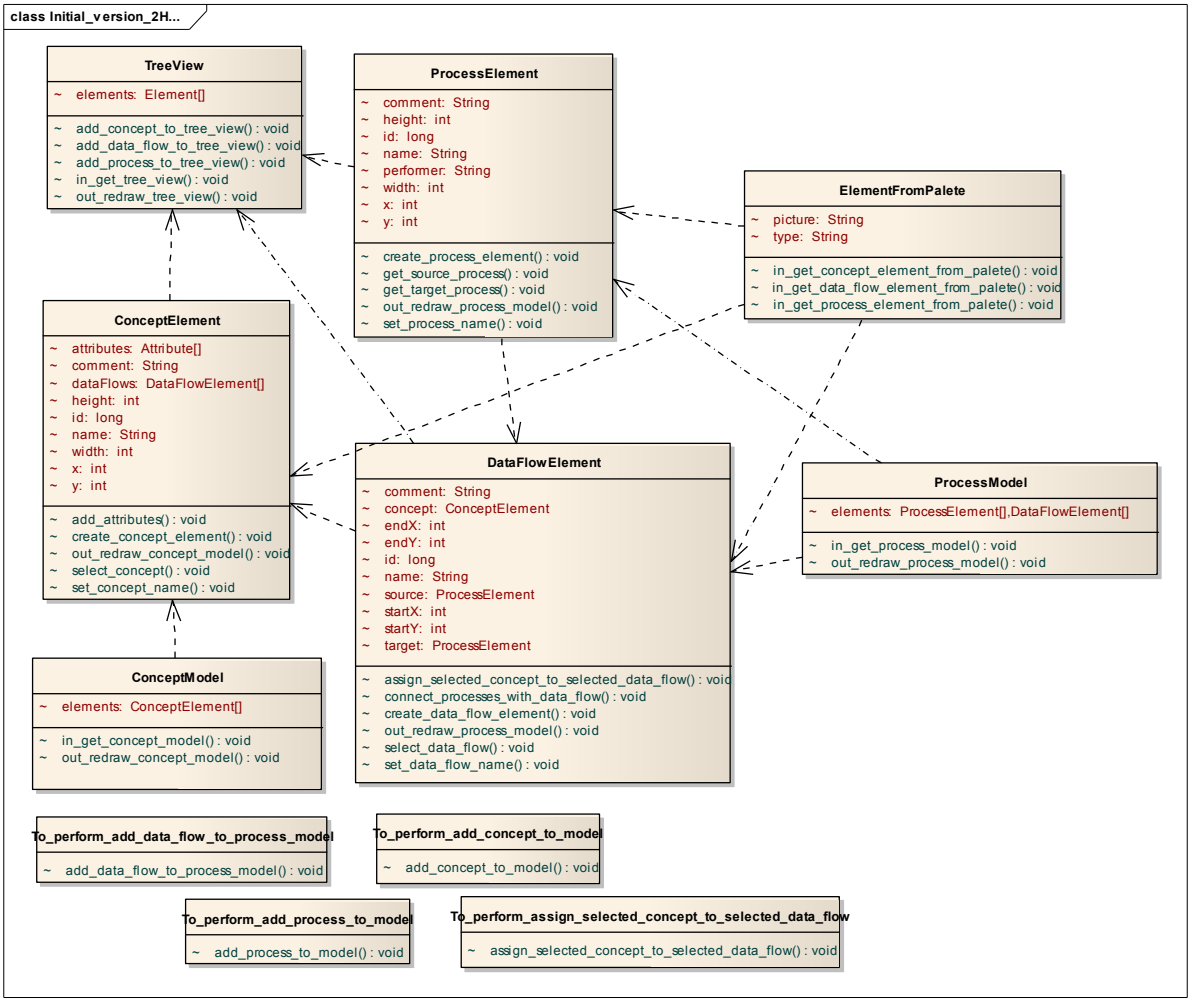


Fig. 2. **Class diagram generated from BrainTool.**

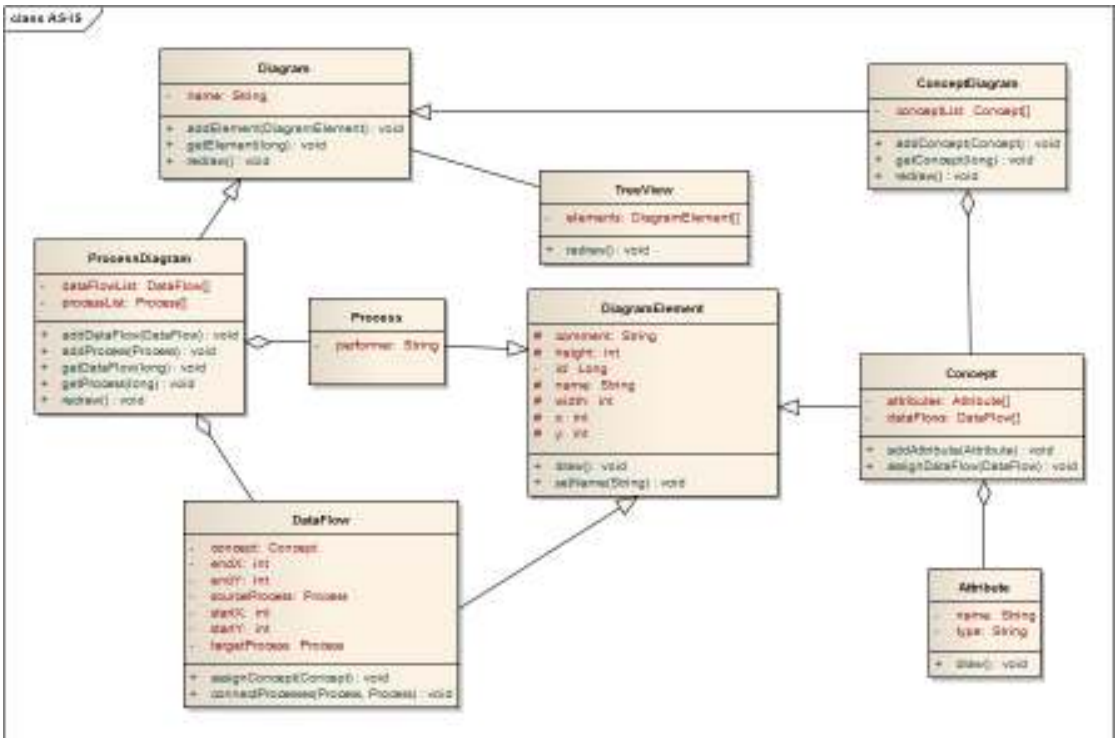


Fig. 3. **Class diagram created during development of BrainTool.**

To perform analysis of the correspondence of the approach supported by BrainTool, three class diagrams are compared each to other. They are the following:

1. The UML class diagram generated by BrainTool. It is shown in Fig. 2.
2. The class structure of BrainTool itself crated by BrainTool's developers manually and can be stated like "AS-IS" model. It is shown in Fig. 3.
3. The UML class diagram representing how should class model look like, which can be stated as "TO-BE" model. This model were obtained based on BrainTool development analysis and contains the improved AS-IS model with purpose to remove logical problems in it. It is shown in Fig. 4.

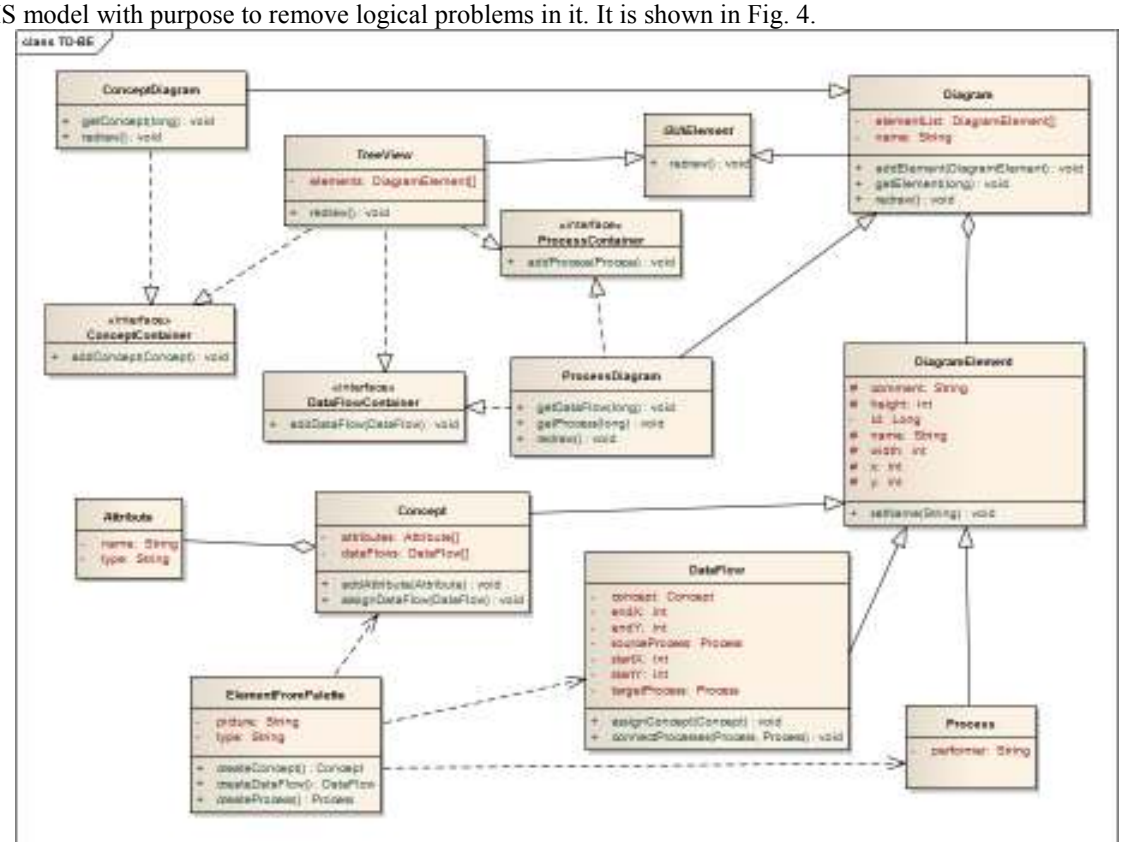


Fig. 4. **TO-BE class diagram.**

To estimate the obtained result, the first two models are compared with TO-BE model. Authors have been selected four comparison criteria describing the core set of the UML class diagram:

1. Classes (how many were generated right classes and interfaces).
2. Associations (how many between obtained associations are right comparison to the TO-BE model).
3. Attributes (how many attributes in each class are the same as in the TO-BE model).
4. Methods (how many methods in each class are the same as in the TO-BE model).

The names, that are in the TO-BE model and in the obtained models may be different, so the comparison were done manually based on obtained elements' semantics. If the meaning of the different named elements is the same, in the comparison process these elements are used as the same elements.

Concerning the classes criteria, authors can claim, that this criteria is good in the Two-Hemisphere approach. All the classes, that is need for the working software were obtained by the automatic transformations in BrainTool. This is expected result, because the class elements is the key elements in the object-oriented software development and reflects the main meaning of the business processes in the software system. However, the current version of BrainTool doesn't allow to generate interface objects for the UML class diagram. These elements might be valuable for the design process result (UML class diagram in our case), but interfaces are not mandatory. As shown on the Fig. 3, the AS-IS diagram doesn't contain interface elements, too. This means, that the resulting code (C# code) isn't so good-looking, but it is working code anyway.

The next criterion is the associations' quality. Authors compared these three models associations. As the result of this analysis, a number of lacks were detected. The main lack is that the generated by BrainTool UML class diagram doesn't contains the generalization association. The ability to create superclasses is quite critical for design process. The next lack is the aggregation correct detection. Though BrainTool provides the possibility to detect aggregation association in the automatic way, the result may be not monosemantic. In this experiment the aggregation associations didn't detect correctly.

The attributes are quite better included in the resulting diagram. Actually, all necessary attributes are including in the BrainTool generated UML class diagram. This means, that all necessary data can be successfully stored in

the software objects. However, the picture is a little differs in all three diagrams. Indeed, because of the lack of generalization association, this attributes may be located in the generated UML class diagram in the different classes, than in the AS-IS or TO-BE models. However, the meaning of the attributes, are they located in the class itself, or in the superclass, is the same.

Analyzing the methods, authors can claim, that the main meaning of the methods in all three diagrams is quite close. Of course, because of the some lacks in the BrainTool's UML class diagram generation process the methods can be located differently, and because of the difference in the design process the names of the methods significantly differs. However, these methods must perform the same actions and they have the same meaning in all presented diagrams.

As the result of the comparison of the automatically generated UML class diagram by BrainTool, manually created during the development process and TO-BE model, authors can claim, that BrainTool is able to generate quite good UML class diagram. This means, that the resulting UML class diagram, produced by BrainTool, have many lacks, however, the data amount and quality stored in the BrainTool's diagram is quite close to data amount and data quality stored in the AS-IS model. In the current state, the BrainTool's UML class diagram can be used in the software development and these results give hope that with the improving of the current BrainTool's lacks, the resulting UML class diagram will be better than the AS-IS diagram. In the current state, BrainTool's result is a little worst than the AS-IS diagram in comparison to the TO-BE diagram. However, the deference is not critical and this diagram (obtained by BrainTool) was obtained in the automatic way, which means, that it is cheaper, than AS-IS model, obtained manually.

However, the necessity to improve the existing transformation process is quite clear. The ideas to imrove the approach to increase the quality of the resulting UML class diagram is given in (Nikiforova et al., 2013). These ideas give the possibility to develop the second version of the BrainTool.

Conclusion

Nowadays, the usage of model transformations has become a widespread practice and tools supporting such transformations have become increasingly popular. The main goal of the research presented in this paper was to verify a tool, which can generate the UML class diagram from the initial presentation of problem domain and to export it to any UML compliant modeling environment supporting the XMI format for model interchange. Authors have been implemented BrainTool that supports creation, editing and validation of the two-hemisphere model and its transformation into the UML class model, where the generated class diagram can further be imported into some UML editor or code generator (Nikiforova et al., 2012a), (Nikiforova et al., 2012b).

The main contributions of the research is a ccomparison of the generated UML class diagram to so called "As-is" and "To-be" diagrams. It approves, that the generated diagram is not worse than the diagram created during the development project. The generated structure has the same classes to support all the processes and conceptual classes defined in the problem domain. And the main benefit is that the structure is received automatically from problem domain presentation. Several distinctions are detected in class' associations. This gives to authors a bases to refine transformations, may be by introducing new elements to the two-hemisphere model.

Within the development process, several new possibilities of two-hemisphere approach were investigated. Authors believe that current transformation rules can be improved in order to reduce the number of limitations currently existing in the two-hemisphere model driven approach, to generate a more precise UML class model and more complete set of the class diagram elements for further using this model for code generation. In turn, several new facilities of code generation directly from the two-hemisphere model were also stated.

Authors' future work will be focused on the implementation of a refined version of BrainTool with respect to the UML class diagram generated from the two-hemisphere model of BrainTool itself and taking into consideration new statements for transformation improvement.

Acknowledgements

We thank the organizers of conference "Engineering for Rural Development" from Technical faculty of Latvia University of Agriculture for permission to use their publication template for our needs. The research presented in the paper is partly supported by Grant of Latvian Council of Science No. 09.1245 "Methods, models and tools for developing and governance of agile information systems".

References

Aalst van der, W.M.P., 2007. Trends in business process analysis: From validation to process mining, International Conference on *Enterprise Information Systems*.

Anderson, J., 1995. Cognitive psychology and its implications, W.H. Freeman and Company, New York.

Balzer, R., 1985. A 15 year perspective on automatic programming, *IEEE Transactions on Software Engineering,* 11 (No 11), pp. 1257-1268.

Kleppe, A., Warmer, J. and Bast, W., 2003. MDA Explained: The Model Driven Architecture – Practise and Promise, Addison-Wesley.

Krogstie, J., 2005. Integrating enterprise and IS development using a model driven approach, 13th International Conference on *Information Systems Development – Advances in Theory, Practice and Education*. Vasilecas O. et al. (Eds). Springer Science+Business media, Inc. pp.43-53.

Loniewski, G., Insfran, E. and S. Abrahao, 2010. A Systematic Review of the Use of Requirements Techniques in Model-Driven Development, 13th Conference, MODELS 2010, *Model Driven Engineering Languages and Systems,* Part II, Oslo, Norway, pp. 213—227.

Mellor, S. J., Scott, K., Uhl, A. and Weise, D., 2004. MDA Distilled. Principles of Model-Driven Architecture, Addison-Wesley.

Nikiforova., O., 2002. General framework for object-oriented software development process In: Scientific Proceedings of Riga Technical University (13), pp. 132–144.

Nikiforova, O. and Kirikova, M., 2004. Two-Hemisphere Model Driven Approach: Engineering Based Software Development, CAiSE 2004 16th International Conference on *Advanced Information Systems Engineering*, June 7-11, Proceedings, pp. 219-233. Riga, Latvia.

Nikiforova, O., Kirikova, M., Pavlova, N., 2006. Two-Hemisphere Driven Approach: Application for Knowledge Modeling, In proceedings of the Seventh IEEE International Baltic Conference on *DB and IS* (BalticDB&IS'2006), O. Vasilecas, J. Eder, A. Caplinskas (Eds.), Vilnius, Lithuania, pp. 244-250

Nikiforova, O., Pavlova, N. and Grigorjev, J., 2008. Several Facilities of Class Diagram Generation from Two-Hemisphere Model In: 23rd International Symposium on *Computer and Information Sciences* (ISCIS 2008). Istanbul, Turkey, October 27-29, [Online]. Available: http://ieeexplore.ieee.org/.

Nikiforova, O. and Pavlova, N., 2009. Foundations on generation of relationships between classes based on initial business knowledge. In: Information Systems Development: Towards a Service Provision Society. Springer US, pp. 289–297.

Nikiforova, O., 2010. Object Interaction as a Central Component of Object-Oriented System Analysis. In: International Conference *Evaluation of Novel Approaches to Software Engineering* (ENASE 2010), Proceedings of the 2nd International Workshop Model Driven Architecture and Modeling Theory Driven Development (MDA&MTDD 2010), Osis J., Nikiforova O. (Eds.), Greece, SciTePress. pp. 3-12.

Nikiforova, O., Pavlova, N., Cernickins, A. and Jakona, T., 2011. Certification of Model-Driven Architecture Tools: Vision and Application, Proceedings of the ICSEA 2011 - The Sixth International Conference on *Software Engineering Advances*. Lavazza L. et al (eds.), IARIA ©, Barcelona, Spain, October 23-29, pp. 393-398.

Nikiforova, O., Pavlova, N., Gusarovs, K., Gorbiks, O., Vorotilovs, J., Zaharovs, A., Umanovskis, D., Sejans, J., 2012a. Development of the Tool for Transformation of the Two-Hemisphere Model to the UML Class Diagram: Technical Solutions and Lessons Learned, Proceedings of the 5th Internation Scientific Conference *Applied Information and Communication Technology*, April 26-27, Jelgava, Latvia, pp. 11-19.

Nikiforova, O., Gusarovs, K., Gorbiks, O., Pavlova, N., 2012b. BrainTool. A Tool for Generation of the UML Class Diagrams. Proceedings of the Seventh International Conference on Software Engineering Advances, Mannaert H. et al. (Eds), IARIA ©, Lisbon, Portugal, November 18-23, 2012, pp. 60-69.

Nikiforova, O., Gusarovs, K., Gorbiks, O., Pavlova N., 2013. Improvement of the Two-Hemisphere Model-Driven Approach for Generation of the UML Class Diagram. Scientific Journal of Riga Technical University: Applied Computer Systems, Grundspenkis J. et al. (Eds), No. 2013/14, Riga Technical University © (in press)

OMG UML Unified Modeling Language Specification. 2011. Retrieved from: http://www.omg.org (last access in January, 2013).

Peyret, H. and Miers, D., 2007. The Shifting Market For Business Process Analysis Tools, *Forester research*.

Rittgen, P., 2010. Quality and perceived usefulness of process models, 25th Symposium On *Applied Computing*. ACM, pp. 65-72.

Stahl, T. and Volter M., 2006. Model-Driven Software Development, Wiley & Sons.

Website of BrainTool. Available at https://webbraintool.org/ (last access in January, 2013).

Introduction of software product and process quality aspects in the study courses of information technologies

Rudite Cevere[1], Sandra Sproge[2]

[1,2] Faculty of Information Technology, Latvia University of Agriculture, Liela str. 2, Jelgava, LV-3001, Latvia
Rudite.Cevere@llu.lv, Sandra.Sproge@llu.lv

Abstract: *Quality assurance aspects related to software product and its development processes is one of the recent topicalities in the sphere of software engineering. Various quality models and approaches to quality evaluation and improvement have been developed, standardised, and documented in any other way. These activities absolutely enable the quality improvement of products in the sphere of information technologies and, thus, simultaneously they enable the quality of business sphere applying the particular information systems. At the same time, it is generally known that the development and introduction of information systems is very labour intensive and expensive process. Hence, it is difficult to find additional time and resources for individual quality evaluation and improvement measures. On such conditions, all the possible preventive measures play an essential role. The research focuses on the possibility to include quality aspects in all the study courses of information technologies, applying quality models of intermediate products, final products, and processes as one of such measures. The impact of individual activities and intermediate results on the quality of final products is exemplified based on the metrics used in the evaluation of quality characteristics. The article outlines activities used for the evaluation of process quality and identified basing on the experience in the evaluation of the software product quality.*

Keywords: software engineering, quality assurance, process quality.

Introduction

Quality assurance aspects related to software product and its development processes have always been topical in the sphere of software engineering. Various quality models and approaches to quality evaluation and improvement have been developed, standardised, and documented in any other way (Thomas et al., 1996). These activities absolutely enable the quality improvement of products in the branch of information technologies, and thus, simultaneously they enable the quality of business sphere applying the particular information systems. The development and introduction of information systems is very time consuming and expensive process. Detection of faults and deficiencies in the end product is very expensive; therefore quality assurance in the course of the development is very important. Analysis and comparison of the quality models (Deissenboeck et al., 2009; Saini et al., 2011; Jamwal, 2010), as well as the development of new ones is still going on (Rawashdeh, Matalkah, 2006). Nevertheless, usage of quality models in practice, and implementation of quality assurance activities are tasks that are not resolved enough.

In such situation, all the possible preventive activities aimed at for decreasing the possibility of fault arising play essential role. The quality improvement activities described in the article are bases on more than ten years' experience in the position of head manager in quality assurance in large IT Company in Latvia. The Company has been ISO 9001 plus TickIt certified since 1999. In accordance to it twice a year the surveillance audit has been carried out. During these audits the progress of active development and maintenance projects has been controlled. The audits have been performed by TickIt certified auditor of the Lloid's Register who continuously performed similar audits in many countries in Europe. Therefore during the audits there was a possibility to get acquainted with actual international trends and requirements in software development. Providing IT business operations of the Company in accordance with requirements of the quality system, over these years the measures that ensured implementation of the quality requirements have been developed and realized in the projects.

In parallel with the implementation of real projects, the proposal arose from this experience how to improve teaching the questions of quality assurance and evaluation in the bachelor's study programmes of information technologies. The traditional way is to teach these topics in separate study courses like *Software Engineering*, *Software Projects Management*, *Software Product Quality*, etc. Teaching such courses is necessary, but is not enough to acquire knowledge about implementation of quality requirements. The experience from the development projects shows that the quality assurance becomes the most effective when these activities become a part of each development task. The approach how to improve the study courses of the information technologies on the bases of this experience is described in the particular research.

Materials and methods

Aim and tasks of the quality system

In accordance with the quality life cycle of software products, the first stage of quality assurance of the end product is related to quality assurance of the development processes. Traditional way of its realization is the development and introduction of the quality system of the company, for example, in accordance with ISO 9001 requirements (Stelzer et al., 1997). The development and maintenance of quality system of an organization has proved itself being a successful decision in the software development (Stevenson, Barnes, 2001). It has also demonstrated long-term experience in IT Company.

Successful quality system has two main prerequisites:

- quality system documentation provides to all staff the minimum substantive guidance as to work in a consistent, disciplined and effective manner;
- regular training is carried out with the staff, ensuring quality assurance measures as a regular part of the work.

Quality management system together with other parts of corporate management systems shall be integrated into a single corporate management system via common elements.

Students have to gain the idea on diversity of quality systems, their significance in the company, advantages and disadvantages of quality systems as well as existing or non-existing risks during their studies when mastering software development and project management requirements, methods and the best practice (for example, software engineering).

Quality assurance activities in software development processes

In a company, the prime business of which is the development of information systems and provision of various information and communication technologies services, the development processes may be defined consistent with recommendation of the ISO 12207 standard (ISO/IEC 12207, 2008). The standard prescribes groups of basic, support, and organisational processes. Basic processes (specification, design, implementation, and testing) are the processes ensuring the development of software. Four support processes – configuration management, change management, review and documentation are the most significant from the quality assurance point of view. The main characteristic of these processes lies in the fact that they are not implemented as sequence of independent operations executed by one performer. Operations necessary for implementation of processes are performed during the basic processes by adding activities required for implementation of support processes to the tasks of basic processes and providing preparation of required documents and records. Therefore, it is necessary to define clearly the interaction of each basic development process with support processes.

The observations of each audit were analysed and the measures were developed for prevention of further origination of similar notes. In such way during several years the necessary quality assurance actions were defined. Their implementation in all projects ensured successful passing of the surveillance audits constantly. The following article contains an analysis of each software product development core process interaction with the support processes, developed on a personal experience.

Requirements specification

Requirements defined during the specification process may be documented in any electronic environment, the content of which and form of documentation shall be coordinated with a customer prior to start of requirements analysis. During the audits the main attention focuses to version control of all developed and received documents, to change control of the deliverables, and to internal reviews of intermediate results and deliverables.

In order to ensure compliance with the specification process quality requirements, the interaction of specification process with four support processes should be implemented. Configuration management, change control and documentation activities are connected to an individual specification activities. Reviews are included as an additional activity in specification process (Table 1).

Table 1

Activities assuring the specification process quality

Specification operation	Related quality assurance activity	Supporting process	Performer	Notes
Collection and summarisation of source documents	Maintenance of source document versions	Configuration management	System analysts	Ensure identification and maintenance of source document versions
Preparation of intermediate results and results	Preparation of transferable documents	Documenting	System analysts	Ensure qualitative form of transferable documents

Specification operation	Related quality assurance activity	Supporting process	Performer	Notes
Preparation of intermediate results and results	Maintenance of transferable document versions	Configuration management	System analysts	Ensure identification of transferable documents
Preparation of intermediate results and results	Maintenance of transferable document changes	Change management	System analysts	Ensure maintenance of change notes of transferable documents
Preparation of intermediate results and results	Reviews of transferable documents and intermediate results	Review process	Group of developers	Ensure quality assessment of transferable documents

Design

The aim of design process is to transform system requirements within design of information system to be developed. Similar to the specification process, the main issues of the audit during the design process are version control of all developed documents, change control of the deliverables, and internal reviews of intermediate results and deliverables. Audit requirements in the design process are usually stronger by require performing design description reviews already in situations when it is not intended as deliverable, and require performing joint reviews.

Thus, during the design, similar to the specification process, it is necessary to ensure the versioning and identification of all design documents, results and intermediate results. Qualitative forms of the design documents, correct change involvement into design description, and requirements traceability to the design elements should be ensured. It is implemented by interaction of the design process with the support processes (Table 2).

Table 2

Activities assuring the design process quality

Designing operation	Related quality assurance activity	Supporting process	Performer	Notes
Preparation of intermediate results and results	Preparation of transferable documents	Documenting	System analysts	Ensure qualitative form of transferable documents
Preparation of intermediate results and results	Maintenance of transferable data item (document) versions	Configuration management	System analysts	Ensure identification of transferable items
Preparation of intermediate results and results	Maintenance of transferable document changes	Change management	System analysts	Ensure maintenance of change notes of transferable documents
Preparation of intermediate results and results	Internal reviews of transferable documents and intermediate results	Review process	Group of developers	Ensure quality of transferable data items (documents) and conformity with requirements
Preparation of intermediate results and results	General reviews of transferable documents and intermediate results	Review process	Group of developers and a customer	Ensure conformity of transferable data items (documents) with requirements

Coding and debugging

In coding process all agreed and approved documents become especially strict configuration management controlled items. Maintenance of the versions, identification and change control should be ensured for all code items. Forms of their realization differ from those of the specifications and design descriptions, but the compliance with these requirements are more stringent. The main auditable issues for coding are requirements traceability, i.e. the compliance of the developed units of the software code with the requirements and design, and compliance of the code with the requirements of a unified coding style. In particular, code reviews are strictly supervised activities. Involvement of the documentation process allows solving the problems with self-documentation of the code and realization of embedded electronic documentation (help). Interaction of the design process with the support processes is described in Table 3.

Table 3

Activities assuring the coding process quality

Coding operation	Related quality assurance activity	Supporting process	Performer	Notes
Preparation of source code data items	Documenting of source code	Documenting	Programmers	Ensure self-documenting of source code
Preparation of source code data items	Maintenance of source code data item versions	Configuration management	Programmers	Ensure identification of software data items
Preparation of executive code data items	Maintenance of executive code data item versions	Configuration management	Programmers	Ensure identification of executive code
Preparation of source code data items	Internal reviews of source code	Review process	Group of developers	Ensure quality of software and conformity with requirements
Preparation of software versions	Reviews of transferable versions	Review process	Group of developers and a customer	Ensure conformity of transferable versions with requirements

Testing

Testing process plays an essential role not only as one of the basic processes of software development but also as a significant activity for quality assurance of final results.

Considering the test document and record diversity, the audit focuses attention to related issues. One of the most popular schemes is an inspection of random problem report for traceability in both directions - from the report to initial requirements and from the report to the program code. In order to ensure these requirements configuration management and change control actions are involved into test documents and records. All necessary reviews should be carried out. Besides the common task of reviews to evaluate current object, reviews in testing should be organized for preparation of management information and for decision making about ending, modification or interruption of testing process. Configuration management has been related with versioning of the test objects, and also with test environment preparation and maintenance. Necessary interaction of the design process with the support processes for testing is described in Table 4.

Table 4

Activities assuring the testing process quality

Testing operation	Related quality assurance activity	Supporting process	Performer	Notes
Preparation of testable objects	Preparation of testing object versions	Configuration management, coding	Testers, system analysts	Ensure timely preparation of testable objects
Preparation of testing documentation	Determination of types for testing notes	Documenting	Testers, system analysts	Ensure preparation of testing records
Preparation of testing documentation	Maintenance of testing document versions	Configuration management	Testers, system analysts	Ensure maintenance of testing records
Preparation of testing notes	Determination of types for testing notes, implementation environment and identification	Configuration management, documenting	Testers, system analysts	Ensure monitoring of testing process information
Preparation of testing notes	Maintenance of testing notes	Change management	Testers, system analysts, programmers	Ensure registration of problem reports and change demands
Preparation of testing documentation and testing notes	Quality assurance of testing documentation and notes	Reviews	Project managers, quality manager, testers	Ensure quality of testing documentation and notes (hence, the entire testing process)
Organisation of testing process	Compliance monitoring of the course of testing performance	Reviews	Project managers, quality manager, testing manager	Ensure adequate procedure of testing process

Activities included in the primary processes provide software quality because maintain an orderly flow of information for all items involved in the development. However, along with the benefits, developers should always keep in mind that quality assurance activities require additional resources. It is unacceptable that this resource becomes unreasonably large. There appears the next aspect of quality assurance, which is related to

quality assessment of the process. The following section describes the approach how the quality model may be used for description of the quality of primary and support processes.

Evaluation of process quality

During the development of the quality system all processes going on in an organisation should be identified and the procedures for their implementation defined. This action results in the development of process descriptions, which form a part of quality system documentation. A description shall be developed for each process irrespective of the fact whether a process is implemented as a sequence of independent operations or as activities added to basic operation of other processes. General description of process includes process inputs, outputs, sequence of the necessary operations resulting in the transformation of input into output as well as resources and control necessary for the process implementation.

General requirements set for process description state that a process shall be described in terms of its purpose and outcomes, and in any process description the set of process outcomes shall be necessary and sufficient to achieve the purpose of the process (ISO/IEC 15504-2, 2003). Results of every process may be necessary to be used for the input of other processes.

Development and description of processes is an indispensable prerequisite for the establishment of quality system. Though, it is insufficient to really implement quality assurance. For quality system's requirements realization in the primary processes of the development the quality activities described above were defined. These activities implement measures necessary for quality assurance of the end product. However, the quality of the development processes in not guarantees that the developed final product - the software will be of good quality. It is necessary yet to take the software product quality and process quality assessment. Thus, it appears another question as to what is considered a high-quality process or product.

Realizing any theory based quality assurance activities, it is necessary to make sure that the potential benefits outweigh the investment of resources. This effect makes urgent need to assess the quality of the process itself.

In ISO 9000 standards there is not defined in detail what process can be thought of as a quality, and how to evaluate the quality of the existing or desired process. These issues are outside the ISO 9000 standard scope. However, the software product quality assessment and quality improvement issues in the field of software development are well developed (Khayami et al., 2009). On the bases on software quality models developed during many years the generalized form of quality model and evaluation techniques are coordinated and approved. They are included in standards of ISO/IEC 9126 and ISO/IEC 14598 series. This approach is further developed in standards of the SQuaRE series (group of ISO/IEC 25000 standards). Approach described in the mentioned standards defines the internal and external quality of software in accordance with quality model of the ISO 9126.

Considering, that during software development it is necessary in different ways to focus on the quality of processes and their outcomes, the approach, described in the current research includes proposal of using unique model for quality description in all cases. The model consists of hierarchically related quality characteristics and subcharacteristics. Therefore the process model also can be defined as hierarchical structure of characteristics and subcharacteristics. (Fig. 1)

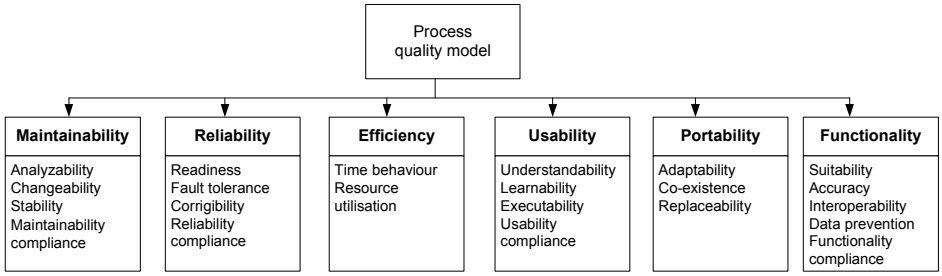


Fig. 1. **Process quality model**.

The same quality characteristics, which are determined for the software product, may also be used for quality evaluation of processes. Certainly, considering that an explanation of the notion may slightly or even greatly differ from the explanation used in the evaluation of products. The largest difference is related with the metrics applied evaluating characteristics or sub-characteristics. For example, a suitability metric "Functional implementation completeness" may be used evaluating functionality of a product. It answers the question "How complete is the implementation according to requirement specifications?" and the method of application is counting the number of missing functions detected in evaluation and comparison with the number of functions described in the requirement specifications. The use of all inputs and acquisition of all outputs will be evaluated instead of specified functions when evaluating the quality of processes.

Particular importance in the successful implementation of the quality system has rational design of the quality assurance processes. Experience gained during the surveillance audits gives a good illustration related to the review process.

Reviews are one of the activities which are very carefully checked during every surveillance audit. A review report, the preparation of which may be quite time consuming, is one of records demanded compulsory. For example, in first years of maintenance of the quality system during a number of audits projects received comments on the lack of the review report in spite that in quality system the review reporting form was designed, that includes all the information required. Therefore, in number of active projects there were carried out observations related to situation with review reports. It showed that, for example, if in a small-scope project one review is performed by 3 employees and the duration of review takes about 1.5 hours, than preparation of a review report according to the form stated by the quality system, may take additional 30-45 minutes for one employee. In addition, for the developers themselves, this neatly presented report is not required, whereas they use their records made during the review. This considerably reduces employees' motivation to perform reviews in general. After a careful analysis of information required by the review report and the information arising as the result of implementation of project tasks, it turns out that only a review register shall be established additionally. Most of information is already recorded in the environment of work flow control, a project plan, e-mail correspondence, minutes of regular meetings, and other records. This example shows necessity for the quality managers to base decisions on a certain understanding of the quality (in the example the assessment was based on the efficiency of the process). For the improvement all employees should master skills of using information of certain type, which is recorded in various places. In any case one should avoid creation a new document with a title of required information, which in fact repeatedly rewrites the existing information. These skills shall be mastered already during the studies.

Using quality model given in Fig. 1 for the quality evaluation of the process, it is desirable to organize decision-making, using the metrics defined by sample of software product quality model (Table 5).

Table 5

Time behaviour metric of the process

Metric name	**Purpose of the metrics**	**Method of application**	**Measurement, formula and data element computations**	**Interpretation of measured value**	**Input to measurement**
Documentation time	What proportion of the time do developers spend preparing the review process documentation?	Observe the reviews of various items carried out in projects. Measure the time it takes to develop the review report.	X = Td / Tr Td = time spent for documentation Tr = total time of reviews	0<= X The smaller the better.	Review report Operation report showing documentation

Results and discussion

In Latvian University of Agriculture, Faculty of Information Technology students are taught in Programming and Computer Science undergraduate programs. In the special courses of the programmes the approach developed by authors of learning the quality assurance issues in-depth gradually are being implemented. The essence of the approach is that in most special courses of information technologies along with special topics should be included the questions about the quality characteristics of the product and about how to change (to increase or reduce) the quality of the end product (Sproge and Cevere, 2012). The reviewed literature has failed to find a description of a similar approach, only a few articles have noted that there should be more attention to the issue of quality learning during the study period (Hilburn and Towhidnejad, 2000). The authors in their research have dealt with issues of how to determine in which courses and to what extent it would be desirable to include software product quality issues.

This article focuses on the acquisition of quality of the basic and support processes of software development. A possible realization of this learning in courses of bachelor's degree program is summarized in Table 6.

Metrics used for software quality characteristics and sub-characteristics may be used for acquisition of software quality indicators, in order to illustrate how particular decisions in code development allows to move towards achievement of quality requirements. For example, testability is one of the sub-characteristics of software maintainability, the influence of which may be determined in the study course by applying metrics given in Table 7. In description of the metric the question tried to answer by help of the metric is formulated, and the quality characteristic to be evaluated is pointed. Besides them, the method of data acquisition and interpretation is described.

Table 6

Quality assurance activities included in primary processes

Software development process	Study Courses	Knowledge and Skills
Specification	*Systems Modelling, Algorithms and structures*	Acquire requirements analysis and modelling tools and techniques or ways necessary for the software development which can also be used to document requirements
	Basics of Software Engineering	Provides knowledge of the organization of specification process and of quality requirements of the specification
	Software Engineering	Acquire configuration management, change control and documentation processes, the practical work contains Reviews of requirements specification
	Programming, Database Technology	In the execution of practical work and course work to follow the minimum requirements of the support process are required (document identification, change control)
Design	*Programming, Database Technology*	The implementation of the practical work should comply with requirements of the configuration management, change management and documentation process, the review of design should be implemented
	Systems Modelling, Algorithms and structures	In practical work review of design is carried out
Coding	*Programming Fundamentals, Programming in Windows environment, Database Application Programming, WWW technologies Programming*	Acquire form and content of the mandatory and recommended comments of source code, maintenance of code changes, and provision of the testability and maintainability of the program. Implement code review (source code data retrieval operator reviews, class libraries reviews, deliverable versions reviews)

Table 7

Testability metrics (ISO 9126, 2003)

Metric name	Purpose of the metrics	Method of application	Measurement, formula and data element computations	Interpretation of measured value	Input to measurement
Completeness of built-in test function	How complete is the built-in test capability.	Count the number of implemented built-in test functions as specified and compare it to the number of built-in test functions in the requirements.	X=A/B A=Number of implemented built-in test function as specified confirmed in review B=Number of built-in test function required	0 <= X <= 1 The closer to 1, the more complete.	A comes from review document B comes from requirements or design document
Autonomy of testability	How independently can the software be tested?	Count the number of dependencies on other systems for testing that has been simulated with stubs and compares it with the total number of test dependencies on other systems.	X=A/B A=Number of dependencies on other systems for testing that have been simulated with stubs B= Total number of test dependencies on other systems	0 <= X <= 1 The closer to 1, the better.	A comes from review document B comes from requirements or design document

The relationship between processes execution and quality models are reflected in Fig. 2. Reviews of different levels are very important during execution of primary processes. It is useful to organize them on the basis of the uniform quality model. The approach, used for quality assessment, which provides metrics purpose formulation in form of a question, makes it possible to prepare checklists which guide each review.

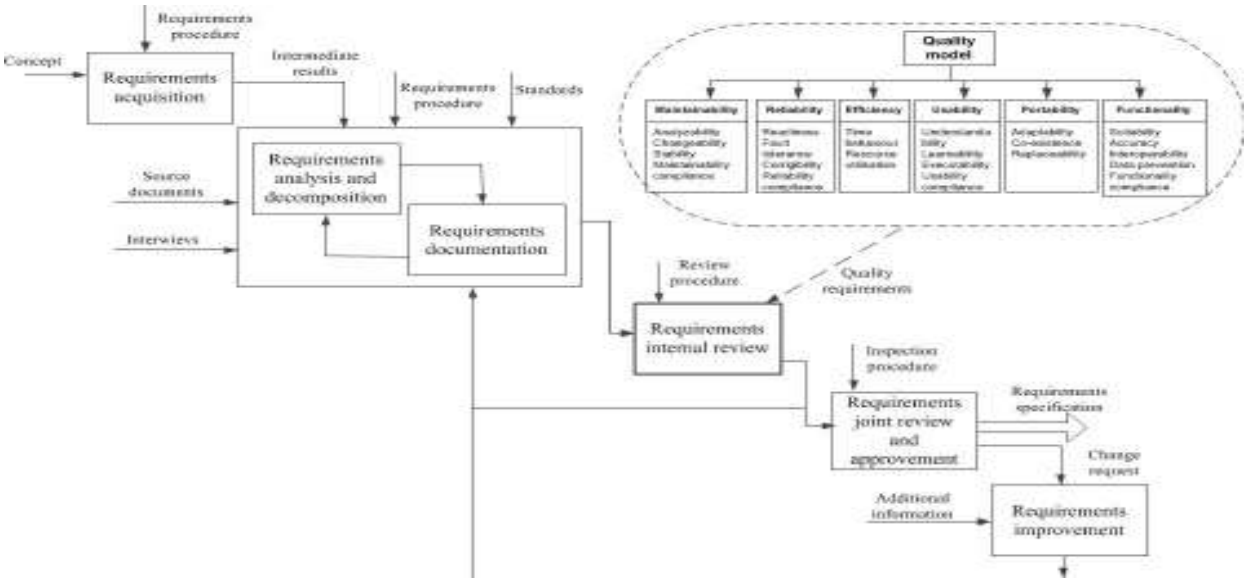


Fig. 2. **Integration of the processes description and the quality model.**

Conclusion

Preventive actions play the most significant role in the sphere of quality improvement, unlike correction of admitted mistakes. During the project preparation and implementation, much attention is being paid to the staff qualification and necessary additional training. Though, it is usually possible to master only programming languages or development environments necessary for the implementation of the particular project due to time restrictions. Therefore, it is essential already during the studies to include quality assurance aspects of corresponding processes or results, their importance and quality improvement activities in the content of field professional specialisation courses when mastering study programmes in information technologies. It is recommended to ground the studying of these quality aspects on a single quality model of a product (end product or intermediate product of software development) and quality model of processes.

During the development and maintenance of software, it is necessary to implement different quality assurance activities, which together create configuration management, change management, documentation, and review processes. These processes are the most important support processes for software development. Simultaneously, the implementation of support process activities shall be included into the structure of every basic process of software development by adding these activities as supplement to the activities of basic process.

To teach basics of quality assurance in training of specialists in information technologies, the study courses shall be developed by analogy and include compulsory and advisable quality activities of each course, outlining their impact on the quality of a product. Quality models of software product and processes may be applied for characterisation of the impact as well as metrics recommended for the evaluation.

The Faculty of Information Technologies of Latvia University of Agriculture has started such improvement of study courses in two bachelor's study programmes "Computer Management and Computer Science" and "Programming" as well as in a master's study programme "Information Technologies".

Acknowledgements

This research has been supported by the European Social Fund within the project "Support for the Implementation of Doctoral Studies at Latvia University of Agriculture". Agreement No. 2009/0180/1DP/1.1.2.1.2/09/IPIA/VIAA/017.

References

Al-Qutaish, R., 2010. Quality Models in Software Engineering Literature: An Analytical and Comparative Study. *Journal of American Science*, 6(3), 166-175.

Deissenboeck, F., Juergens, E., Lochmann, K., Wagner, S., 2009. Software quality models: Purposes, usage scenarios and requirements. In *7th International Workshop on Software Quality* (WoSQ '09). IEEE Computer Society.

Hilburn, T. and Towhidnejad, M., 2000. Software Quality: A Curriculum Postscript? In: *Proceedings of the 31st SIGCSE technical symposium on Computer science education.* New York, 167-171

ISO/IEC 12207:2008. Information technology – Software life cycle processes. International Organization for Standardization.

ISO/IEC 14598-2: 2000. Software engineering -- Product evaluation - Part 2: Planning and Management. International Organization for Standardization.

ISO/IEC 15504-2:2003. Information technology – Process assessment –Part 2: Performing an assessment. International Organization for Standardization.

ISO/IEC 25030:2007. Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Quality Requirements. International Organization for Standardization.

ISO/IEC 9126-1:2001. Software Engineering – Product Quality – Part 1: Quality Model. International Organization for Standardization.

ISO/IEC 9126-3:2003. Software engineering – Product quality – Part 3: Internal metrics. International Organization for Standardization.

Jamwal, D.,2010. Analysis of Software Quality Models for Organizations. *International Journal of Latest Trends in Computing*,1(2), 19-23

Khayami, R., Towhidi, A., Ziarati, K., 2009. The Analytical Comparison of Qualitative Models of Software Systems. *World Applied Sciences Journal,* 6 (1), 1-6.

Rawashdeh, A. and Matalkah, B., 2006. A New Software Quality Model for Evaluating COTS Components. *Journal of Computer Science* **2**(4), 2006, 373-381.

Saini, R, Dubey, R., Rana, A., 2011. Analytical Study of Maintainability Models for Quality Evaluation. *Indian Journal of Computer Science and Engineering*, 2(3), 449-454.

Sproge, S. and Cevere, R., 2012. Position and role of feedback in the software quality life cycle. *Quality Issues and Insights in the 21st Century,* 1, 29-40.

Stelzer, D., Mellis, W., Herzwurm, G., 1997. A critical look at ISO 9000 for software quality management, *Software Quality Journal*, 6, 65–79.

Stevenson, T. and Barnes, F., 2001. Fourteen years of ISO 9000: impact, criticisms, costs and benefits *Business Horizons,* 44(3) (2001), 45-51.

The TickIT Guide. A Guide to Software Quality Management System Construction and Certification to ISO 9001:2000. BSI British Standards.

Thomas, S., Hurley, S., Barnes, D., 1996. Looking for the human factors in software quality management, *Software Engineering: Education and Practice*, 474-480.

Investigation of Linear Genetic Programming Classifiers

Sergejs Provorovs[1], Irina Provorova[1], Arkady Borisov[1,2]

[1] Department of Computer Science, Riga Technical University (RTU), Meza ¼ - 463, Riga, LV-1048, Latvia
[2] Dr.habil.sc.comp., prof.
E-mail: psv@psv.lv

Abstract: *This paper presents a study of linear genetic programming classifiers to solve classification tasks. Usually classification tasks deal with data sets containing a large number of attributes and records, and more than two classes that will be processed using, for example, the obtained classification rules. As a result, by using classical methods to classify a large number of records, the outcome will be a high classification error. In artificial intelligence, genetic programming is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a task defined by a user. It is a specialization of genetic algorithms where each individual is a computer program. It is a machine learning technique used to optimize a population of computer programs according to a fitness landscape determined by ability of a program to perform a given computational task. The linear genetic programming is a direction of evolution algorithm development, whose application areas are not well defined. In recent years different types of genetic programming have emerged. They all follow the basic idea of genetic programming to automatically evolve computer programs. Three basic forms of representation may be distinguished for genetic programs. Besides the traditional tree representations, these are linear and graph representations. Linear genetic programming is a genetic programming variant that evolves sequences of instructions from an imperative programming language or from machine language.*

Keywords: genetic programming, linear genetic programming, data mining, classification task.

Introduction

This work presents an analysis of linear genetic programming capability to solve classification tasks. LGP bases and the principles of its work are considered, and options on the improvement of work of the method are offered. The estimation of LGP algorithm implementation is based on the comparison with the artificial neural networks (ANN) classification result, namely - the classification error.

Background of Linear Genetic Programming

Genetic Programming (GP) is an evolutionary-based search technique which can be applied to many different types of problems. In recent years different types of GP have emerged. They all follow the basic idea of GP to automatically evolve computer programs. Three basic forms of representation may be distinguished for genetic programs. Besides the traditional tree representations, these are linear and graph representations (Brameier and Banzhaf, 2007).

Linear genetic programming (LGP) is a GP version that evolves sequences of instructions from an imperative programming language or from machine language.

In linear genetic programming we can identify two spaces. The phenotype space is a set of mathematical functions, but the genotype space is a set of programs in a certain representation. These programs are composed of elements of the used programming language. LGP programs are represented as instruction sequences that when executed give us the solution to our problem. Each instruction is composed of (Fig. 1):

- A destination register (output) that stores the result;
- One or two registers that store the values used in the operation;
- An operation from the operation set defined.

```
void gp(r)
double r[8];
{ ...
r[0] = r[5] + 52;
r[7] = r[0] - 65;
if (r[2] > 1)
if (r[4] > 5)
r[4] = r[2] * r[1];
r[2] = r[5] + r[4];
r[6] = r[4] * 13;
r[1] = r[3] / 2;
if (r[0] > r[1])
r[3] = r[5] * r[5];
r[7] = r[6] - 2;
r[5] = r[7] + 15;
if (r[1] <= r[6])
r[0] = sin(r[7]);
}
```

Fig. 1. **LGP program example in C notation.**

The LGP uses the following steps to evolve a computer program that predicts the target output from a data file of inputs and outputs (Brameier and Banzhaf, 2007):

- Initializing a population of randomly generated programs;

- Running a selection. In this step four programs are selected from the population randomly. They are compared and based on fitness, two programs are chosen as winners and two as losers;
- Transforming the winner programs. After that two winner programs are copied and transformed probabilistically as follows:
  - parts of the winner programs are exchanged with each other to create two new programs (crossover operation);
  - each of the selection winners are exchanged randomly to create two new programs (mutation operation);
- Replacing the loser programs in the selection with the transformed winner programs. The winners of the selection remain without change;
- Repeating steps two to four until convergence. A program defines the output of the algorithm that simulates the behavior of the problem to an arbitrary degree of accuracy.

In LGP a user-defined number of variable registers, the register set, is made available to a genetic program. Besides the minimal number of input registers required to hold the program inputs before execution, additional registers can be provided in order to facilitate calculations. Normally these so-called calculation registers are initialized with a constant value each time a program is executed on a fitness case. The instruction set defines the particular programming language that is evolved (Brameier M. and Banzhaf W., 2007). Two-operand instructions may either possess two indexed variable (registers) operands or one indexed variable and a constant. One-operand instructions use only register operands. If there cannot be more than one constant per instruction, the percentage of instructions holding a constant is equal to the proportion of constants in programs. This is also the selection probability of a constant operand during the initialization of programs and during mutations. LGP programs are represented as instruction (Table 1) sequences that when executed provides us the solution to our classification problem.

Table 1

General notation of instructions types

Instruction type	General notation
Arithmetic operations	$r_i := r_j + r_k$, $r_i := r_j - r_k$, $r_i := r_j \cdot r_k$, $r_i := r_j / r_k$
Exponential functions	$r_i := r^{r_k}$, $r_i := e^{r_j}$, $r_i := \ln(r_j)$, $r_i := r_j^2$, $r_i := \sqrt{r_j}$
Trigonomic functions	$r_i := sin(r_j)$, $r_i := \cos(r_j)$
Boolean operations	$r_i := r_j \wedge r_k$, $r_i := r_j \vee r_k$, $r_i := \neg r_k$
Conditional branches	$if\ (r_i > r_k)$, $if\ (r_i \geq r_k)$ $if\ (r_i < r_k)$, $if\ (r_i \leq r_k)$ $if\ (r_i = r_k)$

LGP evolutionary operators are very similar to evolutionary operators - *reproduction*, *crossover* and *mutation* used in this research. *Reproduction* involves copying, unchanged, the selected program to the next generation. The *crossover* operator involves two parents. A random sequence of instructions in the first parent is selected and replaced with a random sequence of instructions from the second parent (Fig. 2).

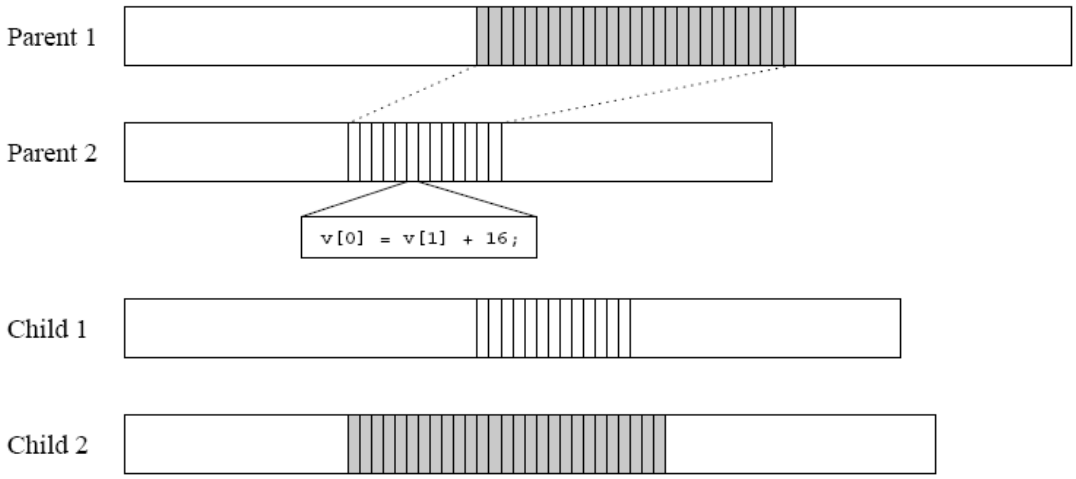


Fig. 2. **Crossover operator in LGP.**

In LGP, for record, crossover points may not be selected inside an instruction and mutations may not exchange an instruction operator for a register. The resulting program is the offspring. If the newly produced program is longer than the maximum length allowed, then an instruction is randomly selected and removed until the program can fit into the maximum length.

Mutation of the imperative representation resulting in the exchange of a register may have a large effect on the functional program structure and on data flow. Even if the absolute program structure is altered only slightly, the effective program may change dramatically. Many instructions preceding the mutated instruction may be deactivated or reactivated. These minimum mutations are possible due to weaker constraints of the functional structure and due to the existence of non-contiguous graph components in linear programs (Brameier and Banzhaf, 2007).

LGP fitness function. The fitness of an individual program is computed by an error function on a set of input-output examples. These so-called fitness cases define the problem that should be solved or approximated by a program. A squared error function penalizes larger errors more heavily than smaller errors. Equation (1) defines a related measure, the mean square error (MSE) (Fogelberg C. and Zhang M., 2005).

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(i_k - o_k)^2. \quad (1)$$

where MSE – fitness function (mean square error);

i_k – the known (input) value of data set;

o_k – the evaluated (output) value of data set.

Detection and deletion of introns. Non-effective code in genetic programs which is referred to as "intron" represents instructions without any influence on the program behavior. Structural *introns* act as protection that reduces the effect of variation on the effective code and also allow variations to remain neutral in terms of fitness change. The imperative program structure in LGP permits structurally no effective instructions to be identified efficiently. It allows the corresponding effective instructions to be extracted from a program during runtime and to be copied to a temporary program buffer once before the fitness of the program is calculated.

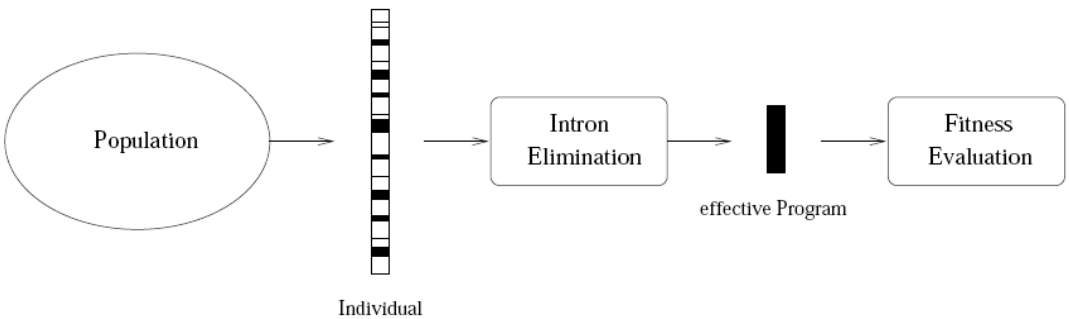


Fig. 3. **Detection and deletion of introns.**

Thereby, the representation of individuals in the population remains unchanged while valuable computation time for non-effective code is saved. No potentially relevant genetic material is lost and intron code may play its role during the evolutionary process. By analogy to the elimination of introns in nature, the linear genetic code is interpreted more efficiently (Brameier M. and Banzhaf W., 2007).

Development of LGP by using "Building blocks"

As is known, the LGP programs are represented as instruction sequences that when executed give us the solution to our problem. These programs are composed of elements (instruction) of the used programming language. This programs call for *chromosomes* and the instructions of program call for *genes* of chromosomes (Provorovs et al., 2012).

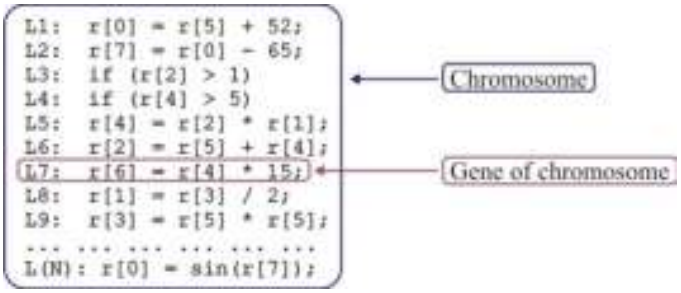


Fig. 4. **Chromosome and genes in LGP**.

The schema of chromosome in LGP is a group of genes (instructions) with particular values (Fig. 5). If chromosomes are "*N*" genes long, and a schema is "*n*" genes long, then *N-n* chromosomes share this schema.

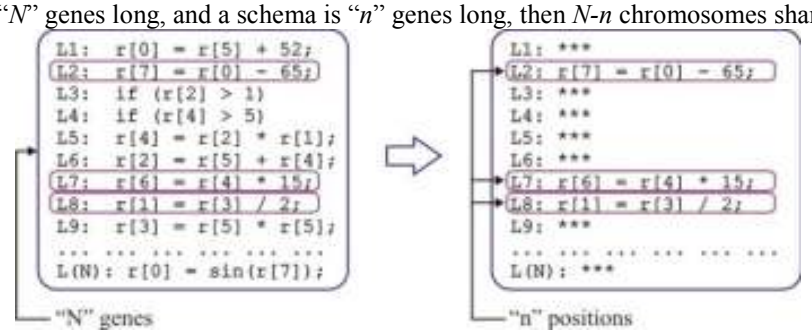


Fig. 5. **Schema in LGP**

The schema is "good" if a chromosome contains it, then the probability that the chromosome has high fitness, is high.
A "Building block" in LGP is a schema: short (few genes) or compact (adjacent genes).

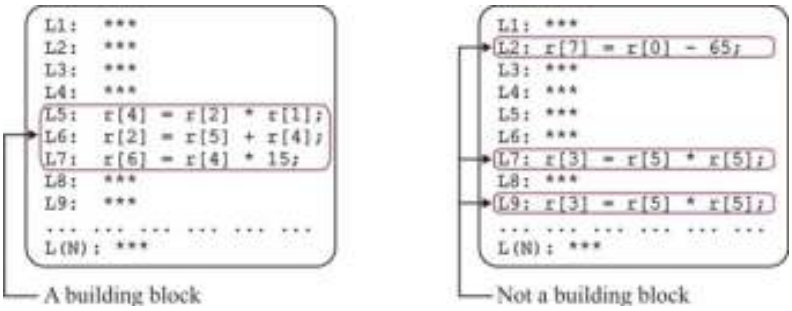


Fig. 6. **Building block in LGP**.

The building block is "good" if, on the average, chromosomes with this building block are "better" than those without it.
The general strategy for applying "building bloks" is: *selection* – favour the proliferation of supposedly good building blocks by cloning good chromosomes more frequently than the bad ones; *recombination* – make it possible for good building blocks from different chromosomes to meet in a new chromosome.

Improvement of LGP by using optimal instruction set

The choice of the instruction set from all available instruction sets (Table 2) is a very important step that effects the LGP classification result – if the choice of instruction set is "successful" then this fact helps to improve the classification result, namely, providing the reduction of classification error, the count of iteration and the count of introns in the programs (Provorovs and Borisovs, 2012).

Table 2

Example of choice of instructions sets

Standard (full) instruction set	Instruction set #1	Instruction set #2

• Arithmetic operations
• Exponential functions
• Trigonomic functions
• Boolean operations
• Conditional branches
	• Arithmetic operations
	• ~~Exponential functions~~
	• Trigonomic functions
	• ~~Boolean operations~~
	• Conditional branches
		• Arithmetic operations
		• ~~Exponential functions~~
		• Trigonomic functions
		  ○ $r_i:=sin(r_j)$
		  ○ ~~$r_i:=cos(r_j)$~~
		• Boolean operations
		• ~~Conditional branches~~

Experiments

The estimation of LGP algorithm implementation is based on the comparison with the artificial neural networks (ANN) classification result, namely - the classification error.

Artificial neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing neurons working in parallel to solve a specific problem. An artificial neural network, often just called a neural network, is a mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data (Provorovs and Borisovs, 2011).

To learn the classification competence of LGP and ANN methods, it was decided to provide a number of experimental series using databases from „UCI Machine Learning Repository" (http://archive.ics.uci.edu/ml) with different number of attributes, classes and records (Table 3).

Table 3

Databases

Database name	Number of attributes	Number of classes	Number of records
Iris	4	3	150
Glass	9	7	214
Heart Statlog	13	2	270
Zoo	17	7	101
Abalone	8	29	4177

Also the experiments were made using "optimal instruction set" and applying "building blocks". The cross-validation was used in each experiment. Special software (RML Technologies, https://www.rmltech.com/store/store.aspx) was used for carrying out these experiments, as well as two programs written were developed additionally.

To make a comparison of working results of LGP and ANN methods, a whole series of experiments were made. Each experiment group used five data bases that differed in their data classification difficulties.

Results and discussion

The results of all experiments produced by the LGP and ANN were evaluated together, grouping by the used data set. As the results of Table 4 show, classification errors depend on the class number – a higher class number gives a higher classification error. Also the attribute number has an influence on the size of classification error.

Table 4

Experimental results of LGP and ANN using multi-classes

Database	Classification error of LGP, %	Classification error of ANN, %
Iris	3.80	4.72
Glass	27.84	30.50
Heart Statlog	22.90	21.26
Zoo	11.46	13.86
Abalone	60.84	45.70

So, to obtain impartial result for LGP and ANN, it was decided to reduce class number by making it the same for each data base.
The results of the experiments with the same class number produced by the LGP and ANN algorithms are evaluated together and summarized in Table 5.

Table 5

Experimental results of LGP and ANN using two classes only

Database	Classification error of LGP, %	Classification error of ANN, %
Iris	2.15	3.27
Glass	11.78	15.82
Heart Statlog	22.36	20.11
Zoo	5.03	7.98
Abalone	28.17	27.72

In general, the results obtained show that in many experiments LGP is better in solving this kind of classification task than ANN.

Conclusion

In most of the experiments performed, the LGP algorithm provides better classification results than the ANN algorithm the classification error is not visibly reduced, it also depends on the classifying database parameters, like the number of classes, records and attributes. By reducing class number to two classes for each database, the classifying error also decreases both for LGP and ANN.
The LGP algorithm classification error also depends on how "successfully" the training and the testing sets are chosen (some of them could not contain any records of one or more dataset classes or the count of represented class records is not proportionate – it complicates the classification process for the class that is represented in minority), and creates the initial population.
Using in LGP the building blocks to solve classification tasks helps to accelerate to create and evolve a "good" program and helps to improve the classification result, namely, providing the reduction of classification error (it was experimentally revealed that the classification error decreases on the average by 7.31%). However, each minor factor (such as the number of records, classes and attributes) could affect the quality of the comparison presented in this paper.

References

Brameier, M., Banzhaf, W., 2007. Linear Genetic Programming – USA: *Springer*, 2007, 322 p.

Frank, A. and Asuncion, A., 2010. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Fogelberg, C. and Zhang, M., 2005. Linear Genetic Programming for Multi-class Object Classification // *Computer Science*, Vol. 38 (2005), pp.369-379.

Provorovs, S. and Borisovs, A., 2011. Use of Linear Genetic Programming and Artificial Neural Network Methods to Solve Classification Task // *Scientific Journal of Riga Technical University*, Datorzinātne, Vol.49 (2011), pp.133.-138.

Provorovs, S. and Borisovs, A., 2012. Using Optimal Instruction Set For Linear Genetic Programming To Solve Classification Task // 16 international conference *"Information Intelligent Systems" journal*, Kharkiv national university of radioelectronics, Kharkiv, Vol.6 (2012), pp.7-8.

Provorovs, S., Provorova, I., Borisovs, A., 2012. Applying Building Block Selection in Linear Genetic Programming to Solve Classification Tasks // Riga Technical University 53rd International Scientific Conference dedicated to the 150th anniversary and The 1 Congress of World Engineers and Riga Polytechnical Institute, *Digest*, 2012, 93 p.

RML Technologies, Discipulus™ 5 Genetic Programming, 2011, [https://www.rmltech.com/store/store.aspx].

The support concept of software development iteration management

Viktors Krasovskis, Ivans Samusenko, Artis Teilans, Oegs Uzga-Rebrovs

Faculty of Engineering, Rezekne University College of Applied Sciences, ITC, Atbrīvošanas aleja 115, Rēzekne, LV-4601, Latvia
viktors.krasovskis@gmail.com

Abstract: *The management process is an important part of almost every software development project. It is a common practice that the same manager is simultaneously involved in multiple projects. It is difficult to focus attention in time to the stages of particular projects, which have increasing problems and risks. Thereby communication between development and management processes is an essential aspect of a successful project. There is a necessity for automatic communication mechanism between development and management processes. Modern development processes are usually divided into multiple iterations, which consist of tasks in order to achieve more transparency and control. The application of ontologies to define the task structure can make it interpretable for data mining. Data can be analyzed using artificial intelligence or modern declarative programming. Thereby, the attributes of current and completed tasks can be analyzed by an applied information system, which will support and/or warn managers to focus on particular projects, basing on the gathered facts, if necessary. Through focusing on the projects, which require management, decision making speeds up. The artefacts of collected development tasks are not wasted and are used to design new standards of tasks and iterations, which are based on successful previous attributes. It can lead to less management. Considering that some projects are almost self-managed and management tasks are applied to the projects, which require management, fewer human resources are necessary. It also positively affects the quality and strictness of management process.*

Keywords: data mining, decision making, iteration, management, ontology, software development, support.

Introduction

Few of significant software development problems are process management, qualitative risk evaluation and well-timed problem solving reaction. Since most processes are managed by humans, there is a high probability for human risk factor. Many software methodologies give common recommendations, but do not describe solutions for particular problems. Respectively almost any methodology is reduced to the fact that development management is grounded on experience from previous projects. This approach is good to solve typical problems, which could be met in the past, but the human factor does not always allow identifying the problem in time.
Since artefact quantity is growing alongside with formalization in almost every middle-sized project, project management will suffer from limited human capabilities extract non-obvious facts from huge amount of data. Thereby projects, which are led using iterative development methodologies, are also affected (Pugh, 2010). Successful run of all iterations is significant to whole project - overall risk factor does not rise. So it is necessary to introduce a semi-automated software development management process, which would allow recognizing problems in time. This paper will describe a support concept of software development iteration management, which is based on Semantic Web and decision making technologies (Allemang and Hendler, 2008).

Materials and methods

One of the ways to improve software developments iterations is to study the development process and to find entry point for improvements. Considering that a support concept of development process iteration management is the main idea of this paper, an abstract iterative development process model needs to be studied. The IDEF0 standard is capable to model this process as shown in Fig. 1 and Fig. 2.

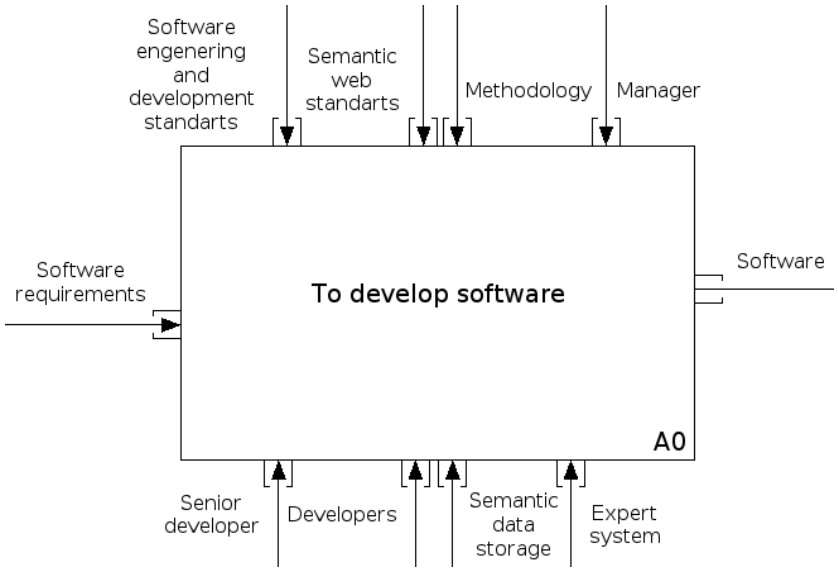


Fig. 1. **Abstract development process model (A0).**

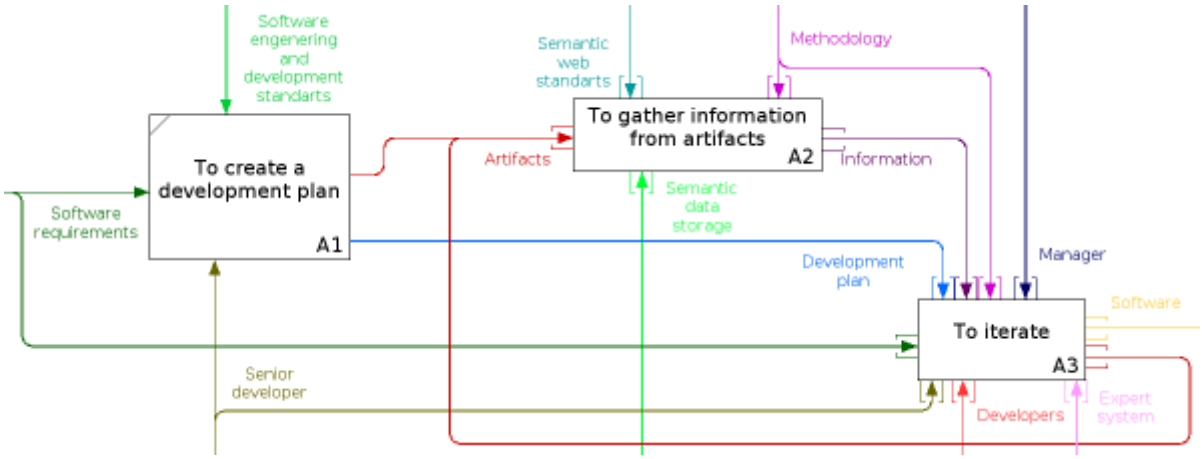


Fig. 2. **Decomposition of A0.**

The process model allows better understanding where it can be improved or changed. One of goals is not only to determine the optimal process modifications for interaction with an expert system for management improvements, but also to determine mechanism and control significance in particular subprocesses.

As mentioned previously an expert system would help to manage the development process. The process itself develops a lot of artifacts, which need to be collected. Later the collected data could be used in the expert system to generate possible solutions for particular problem solving. Data storage in such solution is very significant. It is quite problematical to store artifact data in a relational database because data needs to be prepared before it could be recorded into table. On the other hand also a data extraction problem persists - expensive data mining technologies are needed to use reasoning capabilities. Substantially relational database data mining is based on statistical algorithms, which are good for numerical data processing, but not for other information like triples.

One of the main problems in iteration management is to choose and evaluate right iteration metrics. The data must provide the state of iteration in various contexts e.g. completeness, overall delay, bug quantity dynamics, resource consumption efficiency and overall iteration planning success.

Results and discussion

An abstract iterative development process was chosen to model the interaction between process itself and the expert system (Fig. 2 and Fig. 3). The artifacts gathered from the development plan would be used in information gathering for further iteration control. This subprocess is controlled by the Semantic Web standards and a certain methodology. In this case the main resource is local semantic data storage. The subprocess results the necessary information, which would be used in an expert system to generate a list of possible decisions.

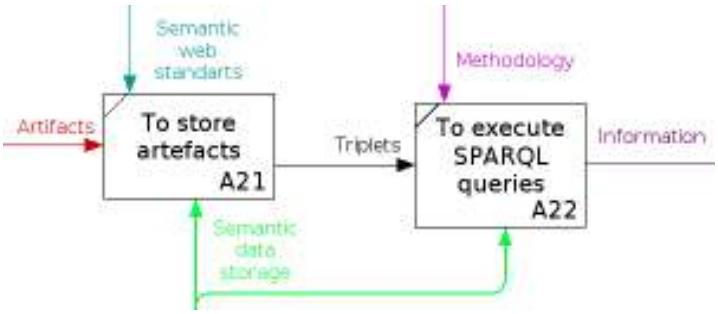


Fig. 3. **Decomposition of A2.**

The iteration execution is also controlled by a certain methodology and by the project's development plan. The third control element is a separate management decision, which is a result of current iteration state rating (Fig.4).

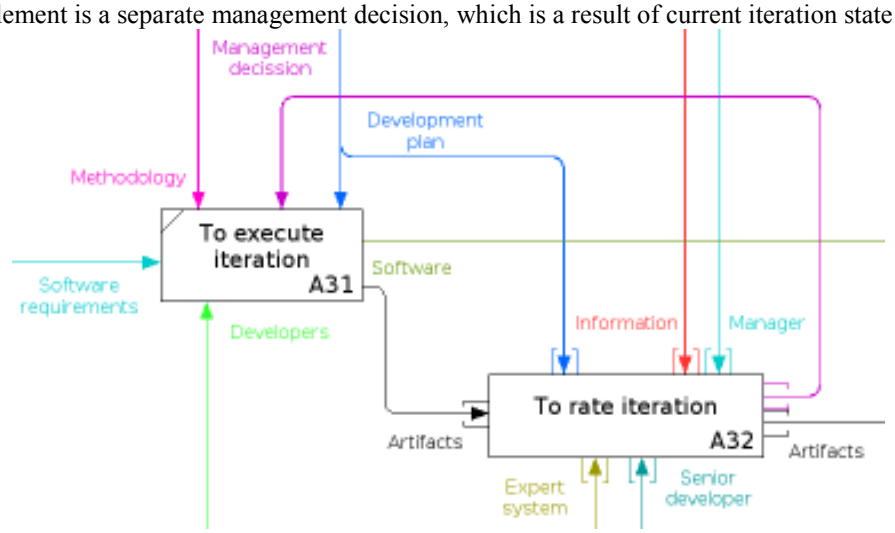


Fig. 4. **Decomposition of A3.**

The iteration rating is a semi-automated subprocess where both experts and an expert system are involved. Live data from current iteration execution is used for input, while control comes from semantic data storage instance. The expert system can contain some predefined rules, which will be used for learning and possible decision generation. If the expert system recognizes a suspicious situation it would start to try generate possible decision, which will be passed to final decision making subprocess where a human manager (also an expert) with help of senior developer will choose the optimal decision if one is present. In opposite case the decision can be made based on personal knowledge and experience (Fig. 5).

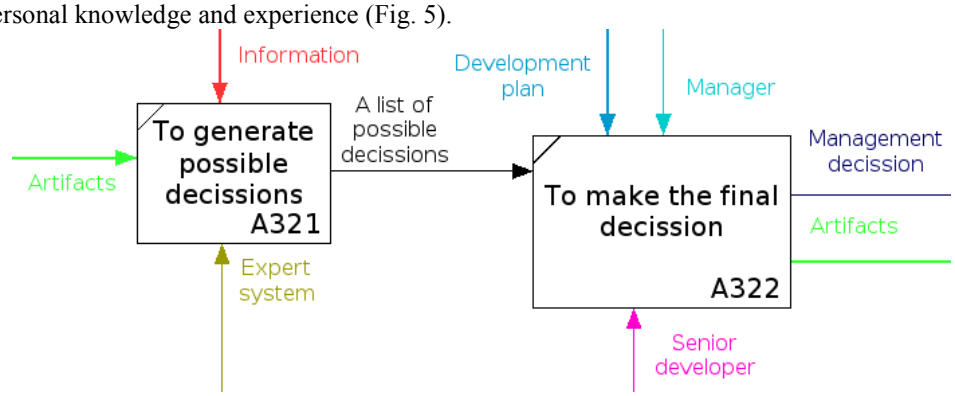


Fig. 5. **Decomposition of A32.**

There are following advantages to use a semantic storage to store data for process automation:

- stores data entities
- reasoning support
- SPARQL (DuCharme, 2011)
- no need to develop complex database schemas
- simple and extended rule definition using RDFS/OWL (Klyne and Carroll, 2002; Brickley et al., 2004)
- many platform-independent storage solutions.

Some current triplestore software products were analyzed for their capabilities (Table 1). They all support SPARQL query language, but only few have RDFS/OWL (Dean and Schreiber, 2004) reverse reasoning support. No all solutions have a built-in Web interface for querying and/or storage management. Only the Virtuoso

universal server has graph storage for triples, which can work in combination with data in relational database. Virtuoso is the only product that supports SQL for additional data management. An essential feature for a triplestore is a built-in SPARQL endpoint, which provides an interface for data querying through other applications or systems.

Table 1

Comparison of triplestore solutions

Product	SPARQL support	RDFS/OWL reasoning support	SQL support	Web interface	Built-in SPARQL endpoint
4store	Yes	No	No	Yes, limited	Yes
ARC2	Yes	No	No	No	Yes
Bigdata	Yes	Yes	No	No	Yes
Mulgara	Yes	No	No	Yes, limited	Yes
Virtuoso	Yes	Yes	Yes	Yes	Yes
RDF::Trine	Yes	No	No	No	No
Redstore	Yes	No	No	Yes, limited	Yes
Redland	Yes	No	No	No	No

Some object class prototype models were developed, which are involved in the development process. These class objects are described using the XML and RDF standards. Almost in every project the atomic piece of every development process is a task (Fig. 6). Any class instance can contain properties, which are essential to subprocess, where they are used.

```
<task:Task rdf:about="http://www.example.org/t3451">
        <task:duration>00:55</task:duration>
        <task:estimationTime>02:00</task:estimationTime>
        <task:complete>70%</task:complete>
        <task:type>1</task:type>
        <task:bugsCaused>1</task:bugsCaused>
        <task:customerMark>6</task:customerMark>
        <task:customerAttention>00:15</task:customerAttention>
        <task:complexity>3</task:complexity>
        <task:priority>1</task:priority>
        <task:iterationSucessInfluence>0</task:iterationSucessInfluence>
        <task:hasCustomer rdf:resource="http://www.example.org/p645"/>
        <task:hasOwner rdf:resource="http://www.example.org/p576"/>
        <task:hasOwner rdf:resource="http://www.example.org/p173"/>
</task:Task>
```

Fig. 6. **Task object example in RDF/XML format.**

Objects can contain other objects as properties e.g. a task can contain multiple owners and customers, but an iteration would contain multiple tasks. The *rdf:about* property describes the object's unique URI, which at the same time is an unique identifier. In turn the *rdf:resource* property indicates the URI of an object in a parent object property. So iterations may contain tasks e.g. using *hasTask* properties (Fig. 7). Thereby objects form relationship with each other or in other words they form a graph.

```
<iteration:Iteration rdf:about="http://www.example.org/i343">
        <iteration:hasTask rdf:resource="http://www.example.org/t3645"/>
        <iteration:hasTask rdf:resource="http://www.example.org/t3451"/>
        <iteration:deadline>2012-11-25 12:00</iteration:deadline>
        <iteration:start>2012-11-21 12:00</iteration:start>
        <iteration:actualStart>2012-11-21 12:00</iteration:actualStart>
        <iteration:actualEnd>2012-11-25 12:00</iteration:actualEnd>
        <iteration:successful>1</iteration:successful>
        <iteration:description>DB model development</iteration:description>
        <iteration:hadManagementIntervention>1</iteration:hadManagementIntervention>
</iteration:Iteration>
```

Fig. 7. **Iteration object example in RDF/XML format.**

Using the RDFS and/or OWL standards it is possible to create schemas, which may describe the rules between object interactions. These rules can be stored separately in a triplestore and be applied to different graphs. The schemas essentially help the reasoning process for gathering non-obvious facts (Fig. 8).

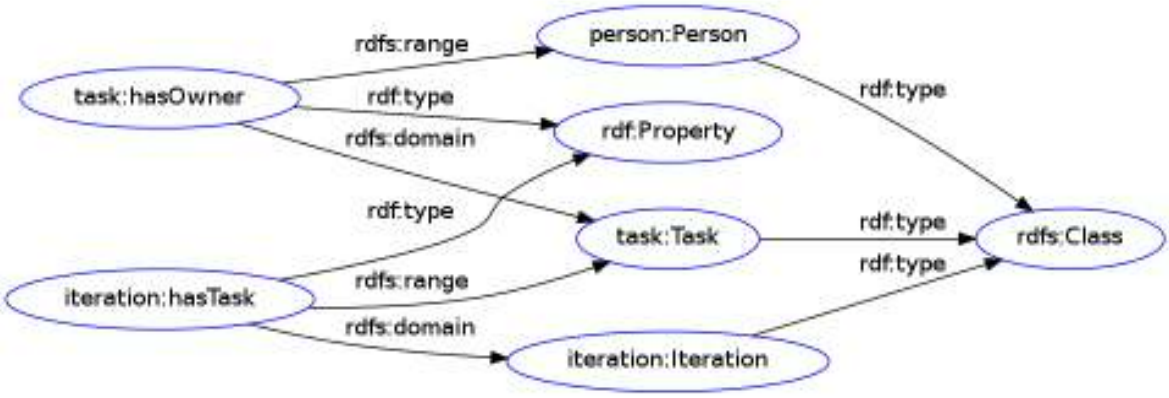


Fig. 8. **RDFS schema example (graph visualization) for the Task and Iteration classes.**

These facts are used in the expert system for decision generation. Since the conditions where decisions have to be made are often variable the expert system has to be dynamic so it will be constantly retrained on new solutions (Shachter and Bhattacharjya, 2010). It is not excluded that some static methods will be used (e.g. correlational evaluation, chi-square) to search for related factors. Classification decision trees, which are intuitive and are used for rules gathering, could be used for information presentation to the end user.

While making decisions in various situations it is possible to face such proposed or being taken solution rating criteria like authenticity and evaluation of the accruing predicted source situation. It is very important to determine the kind of a being taken solution while rating the risk factor of previously mentioned criteria e.g. is it a undeniable key solution or it is a recommendation like solution, which needs further and more deep analysis. Notice that automated decision process unlikely will be used to make decision for solving trivial tasks. On turnover it would be more reasonable to use it to solve tasks from which material, financial values and resources depend. In regret the trust level to suchlike systems is noticeably low due banal skepticism. It is explainable with a fact that only the process owner can take the risk, which can not only impact favorable outcome of current, but also of future processes (Prakash and Shenoy, 1994; Shachter and Bhattacharjya, 2007).

Conclusion

Currently there are many triplestore solutions, which can be used to store and query the stored data. Unfortunately, these features do not provide the product's ability to be used in process automation where reasoning support and advanced management options are essential. At this point Virtuoso server is one of products, which meets most requirements. Future product tests are needed to choose an optimal solution for the practical realization of discussed concept.

Probably a model of decision making, which has to solve a question like task with an answer, is not one of the best suited method to make decisions at all. Respectively the best made decision is a result of systematic and detailed analysis of given problem, during which the problem solution is not so mandatory like the problem understanding point itself. In this case for the individual who makes decisions the final solution is not so important like the possibilities for understanding the fact why the particular solution was selected, but not another one.

The ability to use not always obvious key factors for decision making could significantly improve the trust level of the expert system. Since final decisions are almost always made by a user, he will be in an expert role, who will train the system. The ability to store knowledge in series is preferable in many aspects than application of static algorithms. The dynamically built knowledge base could significantly lower the risk level of taking wrong or not enough suitable decisions as the system will store both right decisions and ones, which could lead to negative consequences.

The future perspective of current research includes the following tasks:

- To develop common ontologies for project methodology, project management tools and expert system support.
- To shape process object model classes closer to real-world process examples.
- To develop an expert system, based on dynamic algorithms for decision generation. The semantic storage's SPARQL endpoint would ensure communication with the expert system not only for data querying, but also to update data e.g. save the decision experience for the future use.

References

Allemang, D. and Hendler, J., 2008. Semantic Web for the Working Ontologist, London, 349 p.

Brickley, D., Guha, R.V., 2004. RDF Vocabulary Description Language 1.0: RDF Schema. Available at: http://www.w3.org/TR/rdf-schema/, 03.12.2012.

Dean, M. and Schreiber, G., 2004. OWL Web Ontology Language Reference. Available at: http://www.w3.org/TR/owl-ref/, 15.12.2012.

DuCharme, B., 2011. Learning SPARQL. O'Reilly Media, 258 p.

Klyne, G. and Carroll, J.J., 2002. Resource Description Framework (RDF): Concepts and Abstract Syntax. Available at: http://www.w3.org/TR/2002/WD-rdf-concepts-20021108/, 23.11.2012.

Prakash, P. and Shenoy, A., 1994. Comparison of Graphical Techniques for Decision Analysis. -: European Journal of Operational Research, Vol. 78, Issue 1, pp. 1-28.

Pugh, K., 2010. Lean-Agile Acceptance Test-Driven Development. Better Software Through Collaboration, Prentice Hall, 345 p.

Shachter, D.R. and Bhattacharjya, D., 2007. Evaluating influence diagrams with decision circuits. - Vancouver: Proceedings of the Twenty Third Conference on Uncertainty in Artificial Intelligence, 8 p.

Shachter, D.R. and Bhattacharjya, D., 2010. Dynamic Programming in Influence Dagrams With Decision Circuits.

Visual programming language for modular algorithms

Rudolfs Opmanis, Rihards Opmanis

Institute of Mathematics and Computer Science University of Latvia, Raina bulvaris 29, Riga, LV-1459, Latvia
rudolfs.opmanis@gmail.com, rihardso@latnet.lv

Abstract: *In this paper we present a way how to describe modular algorithms visually using building block approach. We are using graph drawings to represent algorithm flow. Presented algorithm description language allows to develop algorithms faster, see intermediate results, execute the whole algorithm or just part of it and to run it without setting up complex build environment. In theoretical computer science flow charts are widely used to describe algorithm visually, but it is not possible to execute them. As a result in order to execute it programmer needs to transfer graphical algorithm description into some general textual programming language which is error-prone and time consuming process. Interactive graphical programming environments are available for learning computer programming, but since they are not extendable and provide only limited number of language constructions so they are not applicable for implementing advanced algorithms. We use graph recognition and graph generation algorithms as a use case to demonstrate possible uses of the presented visual programming language.*

Keywords: visual programming language, block diagram, data-flow programming, modular algorithm.

Introduction

Visual computer science alone there are many examples when visual representation of information is preferred over textual. One of most obvious examples is database schema which almost always is shown in form of a ER diagram, but not in textual description. That could be explained by the nature of data, because in database schemas we usually have tables that are linked together with many relationships. That kind of data is hard describe in organized, sequential fashion as text description would demand and it is easier to present drawing to the user and let him/her to decide the order in to traverse the whole data model.
Algorithm that is meant to be executed by a computer can be written in sequential form, mainly because processor of a computer can execute instructions only in determined sequential fashion so we believe that this is the reason why textual programming languages are so popular. Nature of computer algorithm description is sequential enough to be logically described in textual form while still being understandable by programmers.
To implement algorithm in textual programming language in addition to thinking about algorithm itself programmers need to know pretty complicated syntax rules of the language and think about how or organize code so that is easy to test, maintain and extend. While implementing algorithms for graph layout and optical graph recognition we faced several issues.
The second chapter defines type of algorithms that could be developed with proposed visual programming language, but in the third chapter we describe the problem that we faced while working on graph layout and recognition algorithms with textual programming languages. In the fourth chapter we demonstrate our proposed visual programming language.
Nowdays when almost everyone has a personal computer only small part of computer users can write even the simplest “Hello world” program. Classic textual programming languages and programming itself is difficult to learn because it requires skills that many people do not have (Lewis, 1987). It is also a lot harder for software developers to predict all possible scenarios and design complex system that will work and give good results with all possible sets of options that user can provide. So we have to find ways how delegate some part of programming to computer users or users with specific non-programming knowledge (Whitley, 1997). Almost every computer user knows how to work with spread-sheets and how to make simple calculations – visual programming language enables even bigger possibilities for these users to make more complex calculations without learning complex computer programming but using only simple statements to specify some things like where to get input and how to present the results.
Visual programming language is expression tool for visual programming. Visual programming refers to any system that allows the user to specify a program is a two dimensional fashion (Myers, 1990) (Myers, 1986).
Two-dimensional displays for programs, such as flowcharts and even the indenting of block structured programs, have long been known as helpful aids in program understanding (Smith, 1977)
Visual programming languages are designed to make programs from ready-made sub-programs, functions or blocks. Typically visual programming languages create data flow graph – thus why it is called dataflow programming (Bragg, 1994).
Big advantage for visual programming languages is that it is easy to modify program data flow so it is easier to test and develop new version of specific block because it is easy to give input to both of them and then compare results.

Definitions

In this paper we consider only one subclass of algorithms that we call modular. Algorithm is modular algorithm if it can be divided into sub-algorithms (modules) that can be executed in fixed order one after another. For example graph layout algorithm that first positions nodes on the plane, then routes edges and lastly positions all labels next to their owner objects could be considered as modular algorithm, because it can be divided in three sub-algorithms. The first sub-algorithm would be node positioning algorithm, then the second sub-algorithm would be edge routing algorithm, but the last one is label positioning algorithm. Each of these three algorithms uses output from the previous one, but they do not depend on some internal properties of each other. With modular algorithms if we have multiple options for some sub-algorithm we can easily swap them and easily tailor result of the whole algorithm to our needs. For example in previously mentioned graph layout algorithm we could have two implementations of edge router sub-algorithm one would route edges as line segments directly connecting source and target nodes, but the other edge router could route edges with bends to ensure that there are no node-edge overlaps. If we have these two implementations we can use either one of them and graph layout algorithm output also changes depending on selected algorithm.
On the other hand we do not consider insertion sort or binary search algorithms to be modular because they do not have natural division of sub-algorithms that could be executed in specified order.
We call implementation stable if it produces the same result for equal input data. This means that running implemented algorithm with equal data we expect that execution order of all statements will be the same. This is property of implementation is important to guarantee the outcome of algorithm. Most typical causes of non-stable implementations of non-random algorithms are caused by iterations over unsorted collections like sets of maps.

Problem statement

While developing algorithms for image processing and data visualization we observed that we spend huge amount of time for implementing algorithm prototypes rather than coming up with ideas for new solutions. In algorithm implementation phase we organized already implemented algorithms in general structures so that they could be easily reused, but still linking together all necessary sub-algorithms in the implementation of theoretically plausible algorithm required significant amount of time. So it was necessary to come up with solution that would allow us to connect pre-existing sub algorithms faster.
Another issue that we faced was that every next iteration of algorithms prototype was evolution of the previous version with rather minimal changes. Moreover it may as well be that the next version not better than the previous version so it is necessary to run both of them simultaneously on the same set of input data to measure which is better. This problem added requirement to be able to keep several almost identical algorithm implementations in parallel, be able to run them at any time and also to quickly see the differences between any two of them. As developed algorithms got complicated it became obvious that working with long textual documents is not the most convenient solution. After observation that all of our prototyped algorithms are modular it became apparent that visual programming values might solve these problems.
During our research for existing visual programming languages we found out that the best fit is KNIME (Berthold, 2006) programming language which was really close to our needs, because it is easily extensible, it already has well developed tool support based on Eclipse RCP (Eclipse RCP, 2013), however the main obstacle of using it for our needs to quickly prototype various modular algorithms were its unstable behaviour. Unstable in this case means that running program on equivalent input data it can produce different output data, but in KNIME it was possible to run defined algorithm and observe changing order of executed sub algorithms. There were also some other less important reasons that prevented using KNIME modular environment: its rather difficult to integrate algorithm developed in KNIME into classical programming languages and also to use it forces to translate all problems into domain of data mining or statistics, because the basic data transfer object between sub algorithms is table.
Because existing visual programming languages didn't provide all required solutions it was decided to create our own data-flow visual programming language that would allow us to quickly develop prototypes, define stable algorithms, load/save algorithm definitions and represent them in intuitive manner.

Architecture

Design of our programming language is based on block diagrams that are vastly used in theoretical computer science. The problem with blocks in block diagrams is that there are fixed number of block types, but contents of each block is not formalized. Complexity of each block in block diagram can be arbitrary detailed it can be just a simple assignment operator or complex algorithm call. To remove this problem when user can create blocks of arbitrary complexity it was decided to allow only those blocks, which we know how to execute i.e. the ones that have implemented specified interfaces and that are registered in language. We don't fix the set of available blocks, but rather define way how they should be implemented and registered in order to use them. Fig. 3 illustrates single building block in our visual programming language. All required inputs are located at the left side of block representing rectangle, and are documented by tooltip text that is shown when mouse hovers over

each of the inputs. Similarly all produced outputs are positioned on the right side of the block and also provides description in tooltip of each input. Blocks name is shown in the centre of the building block, therefore the user can easily follow algorithm flow and more detailed information is provided only by request.

Fig. 3. **Single building block.**

Blocks are linked together by two types of relations. The first type of relation between block is execution order. The second type of relation between blocks is input-output relation. Input-ouput relation is implemented by the common memory that can be accessed from each block. Every block can announce required values and properties that it produces. For each of these inputs and outputs each block instance provides sets of string keys for its inputs and outputs. If the key of some block's required input property is the same as output key of some other block then it means that the first block uses output of the second block. Visually this key equality is represented as line between respective input and output properties. Fig. 4 illustrates two blocks that are linked together. The rectangular block to the left generates a tree graph, but the block to the right can write graph to the file. Thin solid line between "TreeRandomGeneratorBlock" and "TextFileGraphOutputBlock" shows that "TextFileGraphOutputBlock" will use something that is generated as output in "TreeRandomGeneratorBlock" (in this case generated tree graph) as an input. Thicker dashed line on the other hand shows that "TextFileGraphOutputBlock" will be executed right after "TreeRandomGeneratorBlock". Ellipse in the right side is not a block that will be executed, but shows that "TreeRandomGeneratorBlock" has a constant input value for one of required input properties. In this case it shows that "TreeRandomGeneratorBlock" is configured to generate a random tree graph with 100 vertices.

Fig. 4. **Linked building block system.**

More complicated algorithm that is programmed using our visual programming language can be seen in Fig. 5 where one version of opical graph recognition algorithm is visible. According to our experiments with algorithm prototyping his is typical look of any modular algorithm.

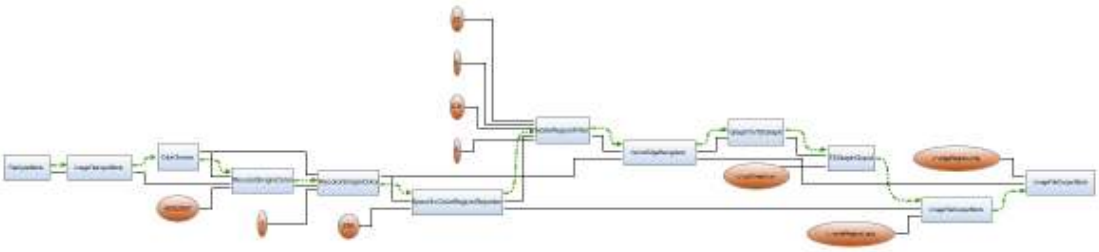


Fig. 5. **More complicated algorithm description.**

Our visual programming language has following features and benefits when compared to traditional textual programming languages:

Interfacing with traditionally written code

When it is necessary to interface visually defined code with some program that is written in conventional programming language such as Java then it simply requires four steps:

1. Read project definition from file or build it with API.
2. Set default values input values
3. Run project
4. Request produced output values from common memory.

Ability to design interactive algorithms faster. By using blocks that presents dialog to the user it is really easy to gather user feedback and adjust future algorithm flow based on entered values. This option might be particularly important to decide if produced results up until now match with desired results. For example when we consider graph layout algorithms it might be useful to allow the user to fine tune automatically positioned node positions before performing automatic edge routing.

Design modular algorithms. Exchanging one block with another requires only two steps. The first step is to create the substitute block, but the second is to reconnect attribute links and control flow path.

Maintain several versions of configurations in parallel and easily compare them. While prototyping it is necessary to maintain some two or more versions of algorithm implementations so that we can compare them, by running on various test data. With textual programming languages it we would need to maintain several classes and use some text file comparison tools to locate the differences between them, but in case of our visual programming language we can perform bipartite matching algorithm between graph objects in both comparable algorithms and visually indicate differences. Our matching allows to draw attention to really significant differences in algorithm structure, but not every single little change in source code like changes in javadoc, additional new line symbols or other coding style improvements.

No need for compilation. Major advantage of our visual programming language over textual programming languages like Java is that our algorithm can be executed without any compilation so compile and run cycle is reduced and prototyping can be done faster.

Implementation details are hidden from the user so developers can use higher abstraction objects. This allows the user to work with higher level concepts and better comprehend the overall picture. This also enables less experienced users to configure complex algorithms, because only general knowledge of each block is necessary.

Easily extendable by implementing general interfaces everyone can extend available set of blocks therefore tailor programming language to any domain where modular algorithms are natural.

Show configuration problems visually. This feature is available in all major IDEs then it shows compilation errors on-the-fly therefore allowing to minimize number of unsuccessful compilation attempts and also visually locating the error. Our visual programming language supports equal functionality. There are certain very simple rules that allows us to show either warning or error message for any block. For example it is considered an error if in control flow there is any block that doesn't have any value or connection associated to its input. This would mean that particular input value is not initialized and it can cause crash in blocs evaluation process.

Building block approach forces to divide implementation in sub blocks and it makes testing easier because it is possible to test each block separately and build algorithms from well tested components.

Our visual programming language allows to look and work with an algorithm from different viewpoints. By default we present algorithm definition in the complete mode (**Fig. 5**) when we illustrate everything that is used to define algorithm: all blocks, execution flow, links connecting produced outputs and required inputs and predefined constant values used for block inputs. It is possible to switch to reduced mode to look separately on control flow (Fig. 6), data flow (Fig. 7) or just block connections (Fig. 8).

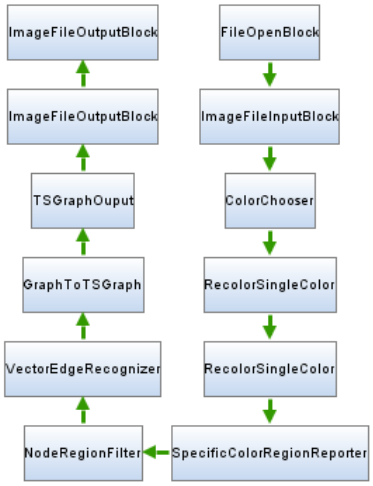


Fig. 6. **Algorithm in reduced mode showing only control flow.**

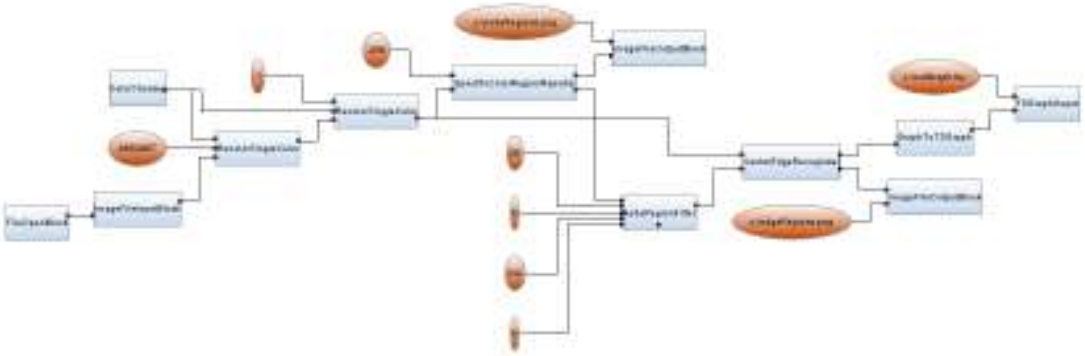


Fig. 7. **Algorithm representation in reduced mode showing data flow.**

Fig. 8. **Algorithm representation in reduced mode showing just block connections.**

Conclusion

We observed that in this paper presented visual programming language significantly increased speed of our algorithm development and allowed to organize various parallel versions of algorithm prototypes while avoiding from copy-paste errors. In future we should do in-depth usability studies with experienced and non-experienced computer users to measure how and if the presented visual programming language reduces development time

Acknowledgements

Supported by ERAF project 2010/0318/2DP/2.1.1.1.0/10/APIA/VIAA/104

References

Berthold, M.R., Cebron, N., Dill, F., Fatta, G.D., Gabriel, T.R., Georg, F., Meinl, T., Ohl, P., Sieb, C., Wiswedel, B., 2006. Knime: The Konstanz Information Miner. Technical Report (http://www.knime.org/).

Bragg, S. and Driskill, C., 1994. Diagrammatic-graphical programming languages and DoD-STD-2167A, Proc. IEEE Autotestcon, Sept. 1994, pp. 211-220.

Myers, B.A., 1986. Visual programming, programming by example, and program visualization: a taxonomy, Proceedings of the SIGCHI conference on Human factors in computing systems, p.59-66, April 13-17, 1986, Boston, Massachusetts, United States

Myers, B.A., 1990. Taxonomies of visual programming and program visualization, Journal of Visual Languages and Computing, v.1 n.1, p.97-123, March, 1990

Lewis, C. and Olson, G.M., 1987. Can Principles of Cognition Lower the Barriers to Programming?, in *Empirical Studies of Programmers*, Vol. 2, Ablex, 1987.

Smith, D.C., 1997. Pygmalion: A Computer Program to Model and Stimulate Creative Thought. Basel, Stuttgart: Birkhauser, 1977.

Whitley, K.N., 1997. Visual programming languages and the empirical evidence for and against. Journal of Visual Languages and Computing, 8 (1997), pp. 9–142.

http://www.eclipse.org/home/categories/rcp.php

INDUSTRIAL APPLICATIONS OF ICT

Electronic document archiving problems and solutions

Toms Leikums

Latvia University of Agriculture, Liela Str. 2, Jelgava, LV-3001, Latvia
Toms_Leikums@programmer.net

Abstract: *Over the past few years both public and private sector organizations gradually switch to electronic document management. Written communication, accounting and other processes related to document management are carried out in digital format. After the expiration of the documents governmental and municipal institutions must submit them to the national archive. The long-established methods that have been used for paper nomenclature and submission of documents are no longer valid when it comes to archival submission of electronically created and signed documents.*
This article deals with Latvian experience of electronic document archiving: solutions and methods of the National Archives of Latvia, up-to-date history of electronic document archiving, norms and regulations requiring the submission of electronic documents to archive. The article also highlights the legislative discrepancies that hinder proper management of the electronic document archiving process. The study also deals with examples from other countries and suggestions for good practice and offers an analysis of several opinions from document management specialists. In conclusion the author of the article proposes solutions for optimization of the whole process of electronic document archiving and electronic document management related suggestions for public sector institutions that would help them to properly prepare for the submission of digital data to state archive.

Keywords: electronic document management, document archiving, public sector document management.

Introduction

Electronic document management has gradually become a standard in every modern organization both in private and public sector. Although a big part of the incoming documentation (letters, submissions) is still in paper format, they are often digitalized and further processed electronically. Governmental and municipal organizations engage in a two-way exchange of electronically signed documents and store them in their document management systems (DMS) in electronic format. In Latvia this particular process has begun later than in Western European countries and therefore governmental organizations have not yet faced the process of submitting all or part of annual documents in electronic format to national archives. The submission to archives is actually the most essential difference between private and public sector. Business enterprises choose to archive documents according to their own needs – to save space, document storage and retrieval costs etc. Public organizations, on their turn, have very strict rules regarding the storage time of specific document types and even subtypes. This period varies from two years up to permanent storage. The documents in national archives are being used in order to receive inquiries, in case of court proceedings, for historical and other various purposes. In every country it is the national archive that overtakes the responsibility of document storage and retrieval after the documents are submitted from governmental organization to archives. Therefore strict and complicated requirements of document submission to archives are understandable. Almost every governmental organization has a special archivist position or even a department for records management that is responsible for document preparation, validity checking, structuring, register maintenance and other preparatory work necessary for successful annual submission of documents to archives. Through several decades the archiving of paper documents had become a perfectly organised and regulated process where all involved parties were well aware of what is expected of them and what to expect from the partners. It is completely different with electronic document archiving – often the archives itself are not certain what data in what format they want to accept. Therefore it is clear that the governmental institutions do not fully comprehend the process of electronic document archiving, too. This article inspects the most important problem questions of electronic document archiving: the necessity of archiving, set of information and metadata to be archived, technological obstacles and archival requirements. In addition, the author also explores the current experience of electronic document archiving in Latvia, offers an analysis of rules and regulations in regards to archiving process and presents scope for future research.

Archiving: basic concepts

In order to comprehend the process of electronic document management, it is helpful to inspect the basic concepts briefly. However, we have to take into account that definitions of particular processes can be significantly different in Eastern and Western document management culture. In Western countries (USA, Canada, Western Europe) document management and records management are two different and separate categories – they can often overlap, yet they are performed separately anyway. Document management is the responsibility of the so called knowledge workers – employees at the institution who create, analyse and process information. They are the ones to create workflows, assign tasks according to specific rank order, create document along with their metadata, and also perform search and control task implementation, if necessary. Whereas it is the records management that is responsible for structuring, archiving and – after a certain period of time – disposing of the data that was created and/or acquired. In Eastern world (former

Soviet Union countries, Eastern Europe) both document management and records management are performed within one process and within a particular institution it is usually an entire department responsible for it. The tasks of this department range from categorizing documents into types and subtypes, creating document nomenclature, defining internal workflows of the institution, ensuring task implementation control, to performing all the functions related to document archiving. In this case employees at the institution have almost no responsibility for document structure and they merely have to create the information and add basic descriptive metadata, along with the participation in already defined workflows (task assignation, implementation, confirmation). In order to comprehend the interaction between document management and records management, we would have to take a look at the united scheme of typical actions in document and records management information system (Fig. 1) by Porter-Roth, presented in his white-paper 'Applying Electronic Records Management in the Document Management Environment: An Integrated Approach'. Even though the scheme describes the particular software product of Xerox – DocuShare – it can be easily replaced with any other document management solution title, and it can be used to describe common processes of document life cycle.

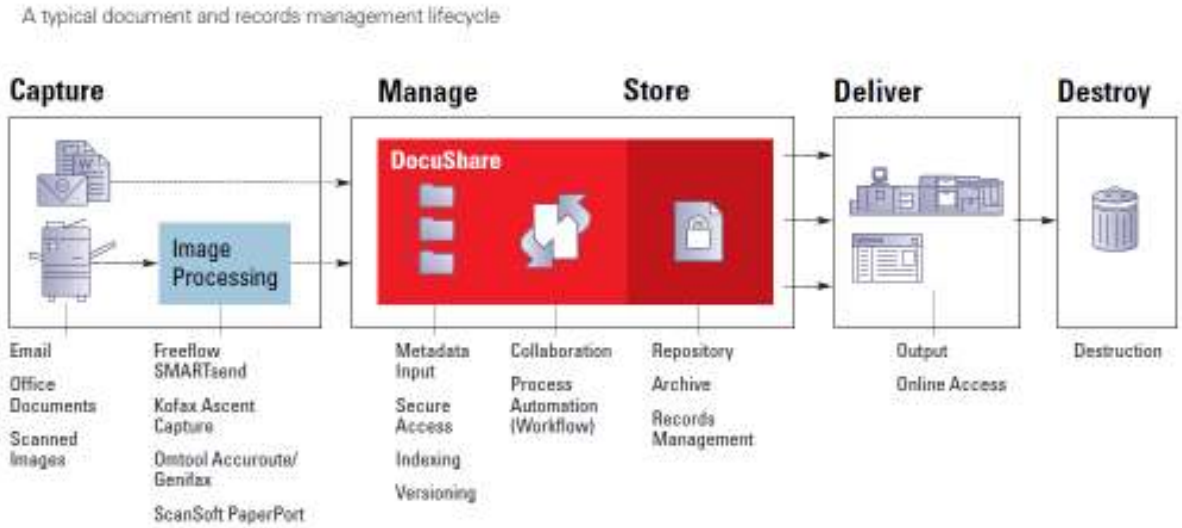


Fig. 1. **Typical document and records management lifecycle.** (Porter-Roth, 2006)

The scheme illustrates the initial input of the information into the system: document can be created in the system or they can be transferred from another information systems, or scanned. Further (sections Manage and Store) the documents are being processed and parallel to it the system already gathers the data needed for archiving. In the context of this article, during the schematic section Deliver electronic records are transferred to an external archive which is usually followed by document disposal from the DMS.

The author of the article has observed that the most appropriate and detailed definition of records management is the one by Australian Standard 4390 – Records Management:

'Records management is the discipline and organizational function of managing records to meet operational business needs, accountability requirements and community expectations. Records management plays many roles within an organization and in the organization's relationships with the world. Thus records management is concerned with the following:

(a) Managing the records continuum, from the design of a recordkeeping system to the end of the records' existence.

(b) Providing a service to meet the needs, and protect the interests, of the organization and its clients.

(c) Capturing complete, accurate, reliable and useable documentation of organizational activity to meet legal, evidential and accountability requirements.

(d) Managing records as an asset and information resource, rather than as a liability.

(e) Promoting efficiency and economy, both in the management of records and in organizational activity as a whole, through sound recordkeeping practices.' (Wardle, 1999).

Other important concepts too look at are archives, electronic record, metadata, storage media, archiving, backup.

Archives: The facilities and supporting resources necessary for the secure retention and maintenance of materials accumulated by an organisation in the conduct of regulatory studies. (Medicines and Healthcare products Regulatory Agency, 2006)

Electronic record: Information recorded in electronic form that requires a computerised system to access or process it. (Medicines and Healthcare products Regulatory Agency, 2006)

One has to take into account that in public sector which is the activity field of the author of the article, only documents officially signed of approved by authorized personnel become electronic records, whereas in private sector the company can choose to treat electronic mail units, meeting notes of employees, document drafts and various other electronic information as electronic records.

The Australian Standard 4390 in reviewing electronic records states that:
'An electronic document becomes an electronic record when it takes part in a business transaction, and is kept to provide evidence of that transaction. To function as a record, however, it must be captured into an electronic recordkeeping system and must incorporate relevant structural and contextual information, as well as content. To be used effectively to manage electronic documents as records, electronic document management systems must incorporate recordkeeping functionality and support the requirements for making and capturing records. (8.4.2) Storage media and related technologies and practices for maintaining electronic records should be specified, designed, operated and maintained so that the evidential elements of the records cannot be altered. (8.7.2) Electronic records should be created and captured into recordkeeping systems in formats that will enable them to be, or at least make it more likely that they will be, available and useable for as long as they are required; i.e. electronic records should be created and maintained in formats that will carry them intact across successive hardware and software platforms'. (Wardle, 1999)
Metadata: Information associated with data that provides context and understanding, i.e. data about data. Most commonly this is data that describes the structure, data elements, inter-relationships and other characteristics of electronic records. (Medicines and Healthcare products Regulatory Agency, 2006)
Storage media: The different materials on which information may be recorded. Examples include paper, photographic film, magnetic media, microforms and optical devices. (Medicines and Healthcare products Regulatory Agency, 2006)
When inspecting electronic document archiving, there is often a question that was not prevailing during the archiving of paper documents – what is the difference between archiving and backup? Often these two processes are not even distinguished and IT specialists assume that they are synonyms. Nevertheless, there is an essential difference between them.

Backup

'The function of backup is to create a duplicate copy of primary data in order to protect that data against loss due to hardware failures, user errors, or data corruption. With backup, a copy of production data is stored in a low-cost format such as tape and often warehoused offsite. Retrieval of historical data from tape backup is a slow, tedious process. Typically backup covers only a limited period of time, for example one to three years.' (Open Text, 2010)

Archiving

'Archiving is a systematic, intentionally designed process of securely storing valuable content in an unalterable and tamperproof form over a long period of time. Long-term storage can be just 7 to 10 years but might well extend to 30 or 90 years and more. The archiving solution secures the accessibility and readability of content during the entire lifecycle and, through replication and distribution, protects it from loss in the event of disaster. The archiving is done in such a way that data can be searched and specific content found and accessed quickly when needed. One reason for retrieving archived data is to meet audit requests or a legal discovery request in the event of pending litigation or the threat of litigation.' (Open Text, 2010)
In addition to that we have to remark that, just as archiving, the backup process is also carried out on a regular basis, yet far more often. Archiving is usually performed annually or semi-annually, whereas backup copies are being created mainly once a day. Therefore backup data are constantly changing, whereas the archived data not only stays the same, but it is not alterable. William Roberts points out that the users are also different. Backup is mainly used by those who create information and current users, whereas archive data are far more often required by persons unrelated to document management: 'Backing up is a regular process required while data is changing on a regular basis. This process typically happens daily, with data retained for a period of months. Restoration is usually within a relatively short period of time, is typically requested by the author, and the prime method of accessing the data is by date. In contrast, the information within a digital archive needs to be accessible many years hence, perhaps by people who were not involved in any way with its generation. For this reason, other related information needs to be stored alongside the original data, and this is usually referred to as metadata.' (Roberts, 2004)

Archiving - why?

Irrespective of the activity type of an organization, the main motive for archiving documents is usually their storage costs. Paper documents are particularly costly for their storage requires space, staff, file cabinets and security level compliant to archiving requirements. Documents created in digital format are much lower and on the whole they could be stored in the current document management system even without submitting them for archiving. However, in this case the main problem would be the running speed of the system. In organizations where tens of thousands of documents are annually processed and where several hundreds of employees are working with them it is an essential decision – what document types to store and for how long? Along with the increasing number of documents the system load is growing and the performance of search queries requires more and more time. Both costs for data storage and creation of backup copies is increasing.
Archiving of electronic documents does not always mean exporting documents from the system and submitting them to an external archive. Many document systems contain implemented mechanisms that enable to create internal archives. Predefined documents from a certain period of time are placed in a separate part of the DMS that is not included in document search queries from the basic interface of the system. As the archived documents cannot be changed, they

also do not have to be regularly backed up. However, this solution requires some disc space, too, therefore many private institutions and all governmental institutions choose to submit documents to archive. Document archiving is approved by specialists of electronic document management as well. Bud Porter-Roth in his study 'Applying Electronic Records Management in the Document Management Environment: An Integrated Approach' analyses potential problems of permanent storage of documents: 'Why not just keep all records all the time? While you may feel safer, and many companies do, by never destroying a record, it is not the best long-term strategy. The most obvious issue is storage space and costs for both paper documents and electronic documents. However, potential legal problems and costs may be a more persuasive reason for not keeping records past their designated retention period. If records are never destroyed, everything is discoverable. Even documents that should have been destroyed will have to be produced. This increases the cost of discovery as there may be double or triple the number of documents that normally would be available had records been destroyed on time.' (Porter-Roth, 2006)
There are also many solutions offering to save on the costs by reducing the volume of electronic documents that have to be stored with the help of compression: 'Electronic archiving can involve concatenating many files, often compressing and encrypting them. Such a practice can reduce storage space requirements and encryption can make the data accessible only to those who have the key for access.' (Ademero, 2012).

Changes in technologies

The main issue while preparing documents for archiving and storing them is always the technological side and changes in standards and software solutions. Over the years many formats of electronic documents have technologically disappeared. It is the task of the organization that maintains the archive to ensure that documents are not too tightly bound to a particular software product or hardware solution. 'Preserving electronic data faces some unique problems. Technological progress is extremely fast in the information-technology industry, and this leads to the hardware and software used for creating the data becoming obsolete fast. Storage media is becoming more compact, and new media using different file formats might not be able to read data created by old media. New versions of software are also introduced regularly to take advantage of technology improvements, and current versions might not be able to read data created by old versions.' (Ademero, 2012)
Long-term storage of electronic documents requires elaborate work on a regular basis. 'After you have constructed your archive, regular testing and maintenance will be required, to ensure that you can still access the information and nothing has been lost. It might even be necessary to convert the archive to new technologies in order to continue its utility. Archiving electronic information requires some thought. If well planned and designed, you will be successful at protecting your organizations information.' (Chester, 2006)
The aforementioned problems do not have one universal solution. William Roberts in his study 'Archiving Electronic Information' (Roberts, 2004) thoroughly describes different ways and approaches how to ensure that data is available for a longer period of time: Hard copy (printing the document and storing the paper), A Museum of Old Technology (maintain old hardware and software in working condition), Emulation (emulation of old software on new hardware/software), Migration (copying of data from one media type to another), Standardisation (an objective of storing data in a standard format is to make use of well-documented and widely available standards to encapsulate the knowledge needed to understand the information in a digital file, rather than depending on proprietary software, where the file format and method of processing are less transparent). William Roberts highlights the migration approach as currently the most popular one. However, it also has many disadvantages – it can come to information discrepancies when another hardware and software is being used for its display. Thus the overall electronic document management in the world is trying to come to unified standards for archiving of electronic documents. Formats of electronic documents in archive are described in next chapter.

Formats of electronic documents in archive

As already stated in the first part of the article, choosing the format for document archive is an essentially important step. On the whole, it is possible to store any file type, thus also the files initially created in or imported into the DMS. These would be, for instance, .doc, .jpg, .xls, .pdf, .txt, .msg and tens of other most frequently used file formats. The main thing is to have the file itself and its respective metadata. However, most electronic document archiving specialists tend to use a particular number of file formats and software able to display them. This is caused by potential patent claims of IT companies for file formats and the software for their display: 'There are many formats and electronic technologies to select for archiving. Those include the ASCII (for the text), TIFF, PDF, JPEG, and XML – not to mention the text processing, the assessments, and other formats. The nature of patent rights of some of these formats leads to the criticism which they cannot be guaranteed to continue for the term.' (El-Bakry, Mohammed, 2009)
A sustainable format for archiving electronic files has to fulfil different requirements. These requirements are listed by Leonard Rosenthol, an expert of electronic document archiving:

1. 'Can convey critical information
2. Can be rendered accurately
    a. 'consistent and predictable'
3. Rich metadata support
    a. Standard schemas such as Dublin Core

   b. Custom schemas
   c. Document providence, version history, audit trails
4. Can incorporate associated 'marginalia'
   a. Notes scrawled in the margins on documents
   b. Comments & markup
5. Digital Signatures for 'tamper-proofing'
6. A definition of how retrieval devices (readers) will behave

I would also add to this list one addition item that I believe is a given, but I want to put it down anyway

7. Based on an open standard (aka non-proprietary).' (Rosenthol, 2012)

Currently in the world there are different approaches towards file format choice. One of the most popular and fast-growing formats is PDF/A format by Adobe. For the promotion of the format the company is using active marketing strategies. Thus PDF/A has already become the standard format for document archiving in many big countries. 'In some markets, using PDF/A isn't just a best practice; it's the law' (Fluckinger, 2009).

PDF/A is an official standard all over the world, for instance, in Great Britain, Australia, Korea, Brazil, Belgium, and many other countries. In the USA as well it is a standard de facto. In spite of the popularity of the format, it also has some opponents, thus signalling that PDF/A format do not lack in shortcomings as well. In his blog Kai Hendry, an expert of electronic document circulation, lists numerous potential problems of PDF/A usage (Hendry, 2012). Though it has to be taken into consideration that the majority of Kai Hendry's listed disadvantages are not current anymore as the list was created in 2009 when the development level of IT and computer networks was lower than it is now. However, the main problems are still intact: PDF with all its varieties is a standard dependent on the developer (Adobe Systems Incorporated); to view PDF files you need separate software (even if it is freeware); PDF files are much larger than other possible archiving solutions (except for TIFF files).

Notwithstanding the fact that PDF/A format has many opponents, its developer is continuing to actively promote its further advance in electronic document archiving field. On a regular basis they organize different events and international workshops in order to improve and develop PDF/A format. More information about such events can be found on the webpage of PDF Association (PDF Association, 2013).

In their materials Adobe Systems Incorporated offer to convert all existing documents in particular DMS to PDF before they are archived (Fig. 2).

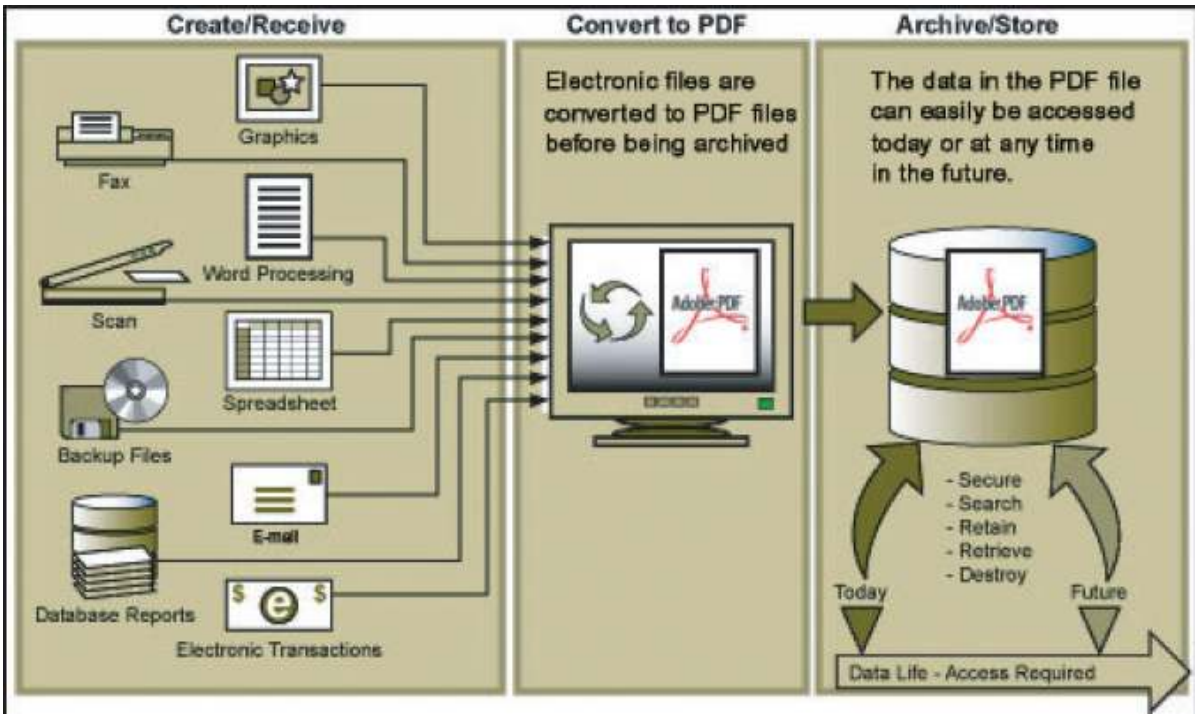


Fig. 2. **Converting and archiving documents from DMS.** (Adobe Systems Incorporated, 2012)

Even though choosing the storing format of documents is a decision made by national archives, without doubt it influences all the institutions in public sector which submit their documents to the archive. In any case the DMS has to be able to export information in one or another format – with or without the manual assistance of an administrator. However, for DMS developers it makes difference if all the file in the system have to be converted into PDF/A or bundled in XML packages, or simply exported as they are, only adding the necessary metadata.

Electronic document archiving – Latvian experience

Discussions about electronic document management in Latvia started when the electronic signature officially became effective back in 2005. However, until the end of 2012 it was only de iure possible to archive electronic documents. In order for governmental institutions to be able to completely switch to electronic document circulation and prepare documents for archiving electronically, they needed officially approved requirements from the National Archives. For several reasons the National Archives had not created unified principles for accepting electronic documents from governmental institutions. An example would be the place of work of the author of the article – Ministry of Agriculture of the Republic of Latvia. Its experience with electronic document archiving started in 2006 when it – as one of the first among Latvian governmental institutions – commenced discussions with the general management of the National Archives of Latvia about the possibilities of electronic document archiving. After several meetings it was clear that archive experts do not have a unified viewpoint on decisions about what data and in what structure they want to receive them from the document management systems of governmental institutions. Instead the Ministry of Agriculture was solicited for offering their own version both about data selection and metadata that the institution could offer the archives, and the means of data submission. In order to store the data from the Ministry of Agriculture the Archives wished to have a fully-fledged and functioning copy of the current DMS of the ministry. Thus additional servers would need to be purchased along with software licenses. Thereby the situation was very uncertain and the ministry was not ready to invest finances into reconstruction of the DMS in order to create a new module for exporting documents and metadata. Moreover, it was not acceptable to create a precedent that one particular ministry would determine a metadata package that would in the future be required by the archive from all the other governmental institutions; for the stored metadata in different DMS are different. As the storage of electronic documents is not compulsory on the state level, the largest part of the public sector organizations decided to continue to submit documents for archiving in paper format as it was considered a cheaper and less resource demanding solution. During the following years IT experts and records managers carefully monitoring all the news from the national archives waiting for some more exact information and detailed requirements for electronic document archiving. Unfortunately the wait and uncertainty lasted for a long time and only by 1 July, 2012 'Procedures for Converting Public Records into Electronic Form' issued by the Cabinet of Ministers of the Republic of Latvia became effective. However, these regulations only covered the proper conversion of original paper documents into electronic form: 'This Regulation prescribes the procedures by which public records shall be converted into electronic form for storage in the electronic environment, the technical requirements for conversion, as well as the procedures by which the obtained records shall be stored and the public records converted into electronic form for storage in the electronic environment shall be destroyed.' (Latvijas Republikas Ministru kabineta noteikumi Nr.143, 2012) In order to start planning the archiving of electronic documents, there were some changes to be done in following regulations: Cabinet of Ministers Regulation No. 747 (Regulation of the Operation of National Archives of Latvia), Cabinet of Ministers Regulation No. 748 (Regulation of Document and Archive Management), and Cabinet of Ministers Regulation No. 749 (Procedures for Document Transfer for Permanent Storage in National Archives of Latvia). The last mentioned regulation were modified and became effective only by 9 November, 2012. The real situation is thus as follows: governmental organizations were able to officially start adapting their DMS only in the last months of 2012 and in an optimal way the first records management year when the electronic documents could be fully submitted to national archives would be 2013. However, the legislative side of electronic document archiving is still not fully regulated and the majority of governmental organizations are hesitant towards the possibility of archiving electronic documents. Even the representatives of the archives admit that some of the aforementioned, newly adopted regulations at some points come into collision with several other legislative acts, for instance, Law on Legal Force of Documents, Electronic Documents Law, Regulations regarding the Manner of Appraisal of Electronic Records, Procedures for the Storage thereof and Transfer to the State Archives for Storage. Currently – by the end of 2013 – a new expert working group is planning to prepare recommendations for altering the legislation in order to make the submission of electronic documents to archives legally fully correct.

As to the regulations that came into force by 9 November, 2012, there is an exact definition of metadata and file types that have to be submitted to archives:

'19. Electronic documents shall be submitted for permanent storage in packages. Each package shall include:

19.1. audit log file in XML format with records about every electronic data and document management system present at the institution wherein user actions have been recorder, if enabled by system functionality;

19.2. file in XML format wherein every submitted file is listed with its name and extension – checksums (hash total) shall be created, using at least two separate algorithms and names of these algorithms;

19.3. structure file of the package in TXT or XML format wherein electronic document checksums (hash total), audit logs, document description files in the package shall be indicated;

19.4. series, subseries, file and document descriptions in XML, XLSX or XLS format;

19.5. electronic documents that shall be submitted for temporary or permanent archival storage.'

(Latvijas Republikas Ministru kabineta noteikumi Nr.143, 2012)

As can be seen, The State Archive of Latvia has chosen to accept .edoc file type documents signed with e-signature and the descriptive and technical information in XML format. Theoretically the popular PDF/A format would also be acceptable, thus potentially decreasing the number of files to archive. However, the author of the article predicts that the

majority of governmental institutions in Latvia will choose – at least in the beginning – to continue submitting documents to archive in paper format or create hybrid files. This would be caused by too rigorous requirements for conversion of original paper documents into electronic documents and their further destruction defined by the particular regulation.

One has to take into account that the vast majority of incoming documents (in many institutions even more than a half) are still in paper format. These are different letters from private persons and organizations, submissions, complaints, contracts, warrants, etc. Thus the only document with legal force is the signed original in paper format. It was already mentioned in this article that there are several discrepancies between electronic document archiving requirements and other legislative acts. In the aforementioned case, if one would have to archive the submission received from a private person in paper format, then it is only the paper document that has some legal force and the archive does not permit different means of digitalization of this document – not even when the authorized persons at the recipient institution make a digitally signed copy of the incoming document or in some other way certify that the newly created document is authentic. In Cabinet of Ministers Regulation No. 143 Annexes 1 and 2 all the required metadata are listed that has to be stored and submitted for a document that has been converted from paper to digital format. Among them are, first of all, the standard metadata as document type, name, registration date, file index in the file nomenclature and the like. These are all the data that can be filled in by records managers or even staff members of the institution on the daily basis. Secondly, there is a vast amount of technical metadata that has to be submitted as well – these cannot be filled in by the records manager because they are out of their scope both because of lacking knowledge and time: user access rights (for the user who performs the conversion of the document from paper to digital format), type of conversion device (for example, flatbed scanner), name and software version of the conversion device, name of the operating system of the computer, software version, applied encryption algorithms (along with the requirement to use at least two of them), reference to the checksum etc. It is certainly possible to include some of these metadata in the program code of the DMS, if the conversion of documents is done by one and the same work station, scanner and software. However, traditionally there is no such function in a document management system and it can be acquired only for additional costs. The conversion of one paper document into digital format can obviously take up to half an hour if archive requirements are followed and all the metadata are filled in. It is not likely that institutions that receive up to several hundreds of documents per day could switch to their electronic archiving even in more distant future.

The author of the article predicts that Latvian governmental institutions will choose to create hybrid files for submission to archives wherein there will be document originals signed both on paper and electronically.

Future research

Records management is a sector that has been busy with documents in paper format for most of its existence and only in last ten years there was a gradual switch to electronic documents. It is likely that in future the proportion of formats will change and processing of information will mainly occur in digital environment. One of the fields for future research is the analysis of already archived electronic documents. Many countries started accepting electronic documents for archiving already at the end of last century therefore there is a chance that many of the formats are not current anymore and soon they even will not be readable. Another possible topic is related to legislative and legal questions that arise when converting a paper document to electronic format. However, it is likely that for a long time the most popular discussion subjects and future research topics will include archiving file formats – if it is optimal to use XML, PDF/A or another format.

Conclusions

Archiving of electronic documents is a topic that gets even more current with every year. During the relatively short period of time since organizations in public and private sector have fully or partially switched to electronic document circulation there is already a vast amount of electronic information that should be archived. It is foreseeable that in the future the number of electronic documents will only increase. Therefore every institution should evaluate its records management policy already now. An important factor is whether the institution does the archiving itself or transfers it to an external archive. The most current question in both cases is the choice of applied technologies and file formats. However, the archiving should not be considered as a burden. Even if archiving of electronic documents is a compulsory process, if properly done, it can improve document management processes in every organization. ‘Archiving documents is a critical function of document management. It can save on costs, improve system performance, and prevent the business from landing in serious legal problems. Special attention must be paid to ensuring readability of the archived electronic data during its lifetime because the data can easily become unreadable when old technology is replaced by new technology.’ (Adamero, 2012).

Many governmental institutions still choose to archive documents in paper format even in cases when the document circulation is already electronic. This is caused by different factors:

Large part of incoming documents is in paper format. As it is the originally signed document that has the legal power, it is understandable why it is required to archive the original. The only way out of this situation would be to create a perfect, safe and standard-compliant environment where the recipient institution could create an electronic copy of the original and verify its authenticity. In this case the original in paper format could be destroyed. However, the number of

requirements for creating such an environment for creating copies is very high and few governmental institutions and their DMS could comply to them without any major changes.

Many outgoing documents are signed on paper in order to respond to correspondents who have submitted their documents in paper format or have requested a reply in paper format. In this case the only solution would be to create two copies of the document and the authorized person would sign one of it on paper and the other one – digitally. The paper version would be sent to the correspondent, whereas the digital version would remain for the institution itself and further archiving. Unfortunately the authorized persons are not always willing to agree to such a solution.

Not all the national archives support the so called hybrid files. As incoming documents in governmental institutions are partially electronic and partially – in paper format, it leads to documents of different types and formats within one nomenclature. For instance, accepting such hybrid files in State Archive of Latvia was accepted only in 2012. One can understand that storing such documents and keeping intact their relation to nomenclature file can cause different problems for the archive because in reality they come from two different environments – physical and digital.

References

Ademero, 2012. 12 Things to Know About Document Archiving. Available at: www.ademero.com/document-management/12-things-to-know-about-document-archiving_article, 22.12.2012.

Adobe Systems Incorporated. PDF as a Standard for Archiving. White Paper. Available at: www.adobe.com/enterprise/pdfs/pdfarchiving.pdf, 06.01.2013.

Australian Standard 4390 – Records Management, 1996.

Chester, B., 2006. Archiving Electronic Files. *AIIM e-Doc magazine*, vol.20.3, pp. 63-64.

El-Bakry, H.M., Mohammed, A.A., 2009. Optimal Document Archiving and Fast Information Retrieval. *International Journal of Electrical and Electronics Engineering*, vol. 6, pp. 356-369.

Fluckinger, D., 2009. PDF/A Takes Root as Digital Archiving Standard. . Available at: http://www.pdfzone.com/c/a/Document-Management/PDFA-Takes-Root-as-Digital-Archiving-Standard, 04.01.2013.

Hendry, K., 2012. PDF/A versus HTML. Available at: http://dabase.com/blog/PDF-A_versus_HTML, 03.01.2013.

Latvijas Republikas Ministru kabineta noteikumi Nr.143. Kārtība, kādā publiskos dokumentus pārvērš elektroniskā formā (Cabinet of Ministers of the Republic of Latvia Regulation No. 143. Procedures for Converting Public Records into Electronic Form). (In Latvian).

Latvijas Republikas Ministru kabineta noteikumi Nr.747. Latvijas Nacionālā arhīva darbības noteikumi (Cabinet of Ministers of the Republic of Latvia Regulation No. 747. Regulation of the Operation of National Archives of Latvia). (In Latvian).

Latvijas Republikas Ministru kabineta noteikumi Nr.748. Dokumentu un arhīvu pārvaldības noteikumi (Cabinet of Ministers of the Republic of Latvia Regulation No. 748. Regulation of Document and Archive Management). (In Latvian).

Latvijas Republikas Ministru kabineta noteikumi Nr.749. Kārtība, kādā nodod dokumentus pastāvīgā valsts glabāšanā Latvijas Nacionālajā arhīvā (Cabinet of Ministers of the Republic of Latvia Regulation No. 749. Procedures for Document Transfer for Permanent Storage in National Archives of Latvia). (In Latvian).

Medicines and Healthcare products Regulatory Agency, 2006. Good Laboratory Practice. Guidance on Archiving. London, United Kingdom, pp. 3.

Open Text, 2010. Why Archiving Matters. How to Manage Your Enterprise Content with Open Text Archive Server. White paper. Available at: www.inge-com.fr/documentation/opentext/ecm/otc_whitepaper_lifecycle_archive_0410.pdf, 20.12.2012.

PDF Association. Events. Available at: http://www.pdfa.org/events, 15.01.2013.

Porter-Roth, B., 2006. Applying Electronic Records Management in the Document Management Environment: An Integrated Approach. White paper. California, USA, pp. 4-5.

Roberts, W., 2004. Archiving Electronic Information. Oxon, United Kingdom, pp. 3-7.

Rosenthol, L., 2012. PDF/A vs. HTML for (long-term) 'born-digital' document archiving. Available at: http://acroeng.adobe.com/leonardr/PDFA_vs_HTML.html, 02.01.2013.

Wardle, D., 1999. ACT Archives Project. Issues and Options Paper. Canberra, Australia, pp. 11-17.

Global search algorithm for the problem of optimization cascade of extractors

Anna Pevneva

National Mineral Resources University 199106 21 line-3, Saint-Petersburg, Russia
pevnevaa@inbox.ru

Abstract: *The article proposes the theoretical scheme of extreme computer simulation to determine the global maximum of the objective function. This is represented in the form of a step-by-step. As the strategy of building of experiment is selected a random search algorithm with directional cone. Modification of this algorithm is based on the principle of sifting "unpromising" trajectory search. Result of this modification can significantly reduce the amount of calculations. In addition, estimates are given for the empirical parameters of random search and some advice on their choice. As well as a comparative table of search results is displayed for different values of parameters. The basis of computer simulation is mathematical model of the cascade of extractors with recycle. It is the system of algebraic equations for the concentrations of substances extracted and balance equations. The criterion adopted to optimize present value of the plant. This value is linearly dependent on the cost of reagents and concentrations of extracted substances. After determining the approximate global optimization is performed by local refinement of this method (used known simplex method). The results of this computational problem clearly demonstrated the need for a global approach in the design of extreme experiment..*

Keywords: global maximum, random search algorithm, directional cone

Introduction

Classification of methods of the theory of global optimization contains two classes. The first includes some methods, based on information about the mathematical properties of the objective function, such as continuity and differentiability. The second area involves iterative algorithms - a series of rules of selecting the best values for each step. These methods mostly use a probabilistic approach in the search. It means, the best value of the objective function is defined with a certain probability. Quite often, these methods are called "heuristics" Markov random search algorithm is a heuristic. This is a Markov chain in which the probability distribution *(k +1)*-th point x_{k+1} depends on the location of the previous point *X* and the objective function *F (x.)*
Markov global random search has common limitations of heuristic methods. First, the convergence of the algorithm, according to various estimates, can be very slow. Second, class of problems for which they are effective, is not clearly defined, and there are no clear recommendations for choosing a large number of parameters that affects the computational efficiency. Despite these properties, random search algorithms are popular among the users because of the relatively simple structure and the possibility to adapt the general scheme for a specific task.

Problem determination and solution algorithm

Global optimization of function *F (x) means* to find the minimum (or maximum) value of the function and the point *X* into the set of optimization, where this value is reached (Zabinsky, 1991; Weise, 2009; Neumaier, 2000).
A global maximum $x^* \in X$ of one (objective) function $F : X \rightarrow R$ is an input element with

$$F(x^*) \geq f(x) \ \forall x \in X.$$

A global minimum $x^* \in X$ of one (objective) function $F : X \rightarrow R$ is an input element with

$$F(x^*) \leq F(x) \ \forall x \in X$$

Below is given step by step presentation of the algorithm of how to find global optimization of function *F*, which given analytically.

1. Randomly choose a point $x^{(i)} \in X$ .. This is the starting point of the trajectory of the search.
2. From this point, according to the formula $x_j^{(i))} = x_i + a \cdot \xi \quad j = 1...m$, where a-set value, ξ random vector uniformly distributed on the interval [0, 1] is determined by *m* random points,
3. For each condition $x_j^{(i)} \in X$ is tested, if it is not satisfied, the point is excluded from the search.
4. Calculated value $Y^{(i)}_j$ of the objective function *F* at the points, choose a point *x**, corresponding to the optimum value of the objective function *F*.
5. From this point, is still m random points $x_k^{(i)} = x^* + a\xi \quad k = 1...m$, those that belong within the hyper-cone $K^{(i)}$ with the axis-hyperlines ($x^{(i)}$, *x **) and data-angle Ψ are selected, that is, the condition

$$\frac{\sum_{l=1}^{n} a_l b_l}{\sqrt{\sum_{l=1}^{n} (a_l)^2}\sqrt{\sum_{l=1}^{n} (b_l)^2}} \le \arccos\frac{\psi}{2}$$

where
$$a_l = (x^*)_l - (x^{(i)})_l$$
$$b_l = (x_k^{(i)})_l - (x^{(i)})_l$$

Here round brackets indicate that the considered l-th coordinate of this point.

6. In the set of selected points one $x_{j+1}^{(i)}$ to be chosen in which the objective function is optimal.
7. Check the condition: if for three consecutive points x_{j-2}, x_{j-1}, x_j holds

$$F(x_{j-2}) > F(x_{j-1}) < F(x_j) \tag{2}$$

then the point x_{j-1} is marked as candidate for minimum (or maximum, if opposite inequalities is right) on this path ends, go to step 1. If all values in a row increasing, opening angle hyper cone halved, go to step 4 with replacing $x^* = x_{j+1}^{(i)}$.

If not satisfied neither one nor the other, then go to step 4 without changing the angle of the disclosure. If the motion on this trajectory is more than *p*, then go to step 1.

Ways to increase the efficiency of the algorithm mostly limited to the successful selection of distributions. High empirical work is a source of instability of the method. The following is a modification that reduces the computational cost of the implementation of the algorithm. The idea is to reduce the number of paths through screening of "close" and "similar." This screening is performed in the search for more specific rule (Ananchenko, 2004).

Thus, the sequence $\left\{Y_k^{*(i)}\right\}_{k=1}^{p}$ is called a complete trajectory of the search $T^{(i)}(p)$ starting from the point $\vec{x}^{(i)}$. Points $\vec{x}_k^{*(i)}$ called nodes trajectory. *P* is the number of nodes for up trajectory. Sequence $\left\{\vec{x}_k^{*(i)}\right\}_{k=1}^{p}$ called the projection of the trajectory on the set *X*.

It is expedient to use the empirical screening procedure of unpromising paths during construction of the search. The principle of this procedure is as follows:

At step *m* (empirical parameter) the particle trajectories $T^{(i)}(p_1)$, *i = 1 ... m* are compared. If for some subsequence $T^{(i)}(p_1)$ and $T^{(j)}(p_1)$ executed

$$\left|Y_k^{*(i)} - Y_k^{*(j)}\right| < r_y \tag{3}$$

And for their projections made

$$\left|\vec{x}_k^{*(i)} - \vec{x}_k^{*(j)}\right| < r_x \tag{4}$$

Then, if

$$Y_k^{*(i)} < Y_k^{*(j)} \tag{5}$$

the trajectory $T^{(i)}(p_1)$ is eliminated as unpromising. Search to be proceeded for the trajectory $T^{(j)}(p_1)$. If the opposite inequalities are right then trajectory $T^{(j)}(p_1)$ is eliminated.

If only the inequality (3) is right, then the search path $T^{(j)}(p_1)$ for the next step increases the parameter a, which determines the magnitude of the next neighborhood node x_k and the opening angle of the cone. If (3) holds for the following *n1* steps, the path is eliminated as unpromising.

Methodology of the global random search algorithm allows creating of the algorithm of sequential design of experiments to determine the global optimization of the response surface. If the investigator has the means of simulation, calculation of objective function can be represented by conducting numerical experiments.

Equations of mathematical model

As the example, the problem of determining the maximum present value from cascade of four extractors with recycle are considered (Fig.1) . The choice of this problem is explained by rather complex way to calculate the target function. It is required not only calculate the value of the criterion (the phrase is simple), but solve the system of nonlinear algebraic equations at first (Lebedeva and Kholodnov, 2004; Kholodnov and Sirenek, 1997).

.

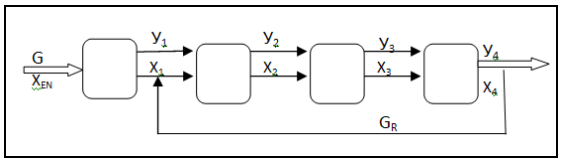


Fig. 1. **Scheme of cascade of four extractors with recycle.**

The mathematical model is a system of algebraic equations for the concentrations of substances extracted.

$$G \cdot x_{En} + G_R \cdot x_4 = (G + G_R) \cdot x_1$$
$$(G + G_R) \cdot x_1 - W_2 \cdot y_2 = (G + G_R) \cdot x_2$$
$$y_2 = f(x_2)$$
$$(G + G_R) \cdot x_2 - W_3 \cdot y_3 = (G + G_R) \cdot x_3$$
$$y_3 = f(x_3)$$
$$(G + G_R) \cdot x_3 - W_4 \cdot y_4 = (G + G_R) \cdot x_4$$
$$y_4 = f(x_4)$$

Where:

G, G_R - consumptions of input stream and recycle, kmol/hr,
x_{En} - the concentration of a substance extracted from the input stream, mole fractions,
x_i, y_i, i=1..4 - concentrations of substances extracted in the extract and raffinate streams, mole fractions,
W_2, W_3, W_4 - consumptions of extragents, kmol/hr,
$f(x_i)$, i = 2 .. 4, define balance:

$$f(x_i) = \begin{cases} 2,5 \cdot x_i + 3,7 \cdot x_i^2 + 11,3 \cdot x_i^3 & if \;\; x_i \le 0,1 \\ 3,9 \cdot x_i + 2,6 \cdot x_i^2 + 7,4 \cdot x_i^3 & if \;\; x_i > 0,1 \end{cases}$$

where numeric coefficients are defined from empyreans. These numerical coefficients are close to those of full-scale experiment.

Optimization criterion is revenue from the installation

$$CR = Z \cdot G\,(x_{En} - x_4) - as \cdot (W_2 + W_3 + W_4).$$

This problem is to find the global optimum is interesting, first of all, because the optimization set is not simply connected. Due to the complexity of the chemical composition of the mixture it is impossible to use regular methods.
Where
Z - factor taking into account the cost of recoverable material.
as - coefficient, which takes into account the cost of the solvent
Obviously, these coefficients have a significant impact on the value of CR. But there are no any clear recommendations for choosing values of Z and as. In the model computational problem objective is the study of the method.
Variable value is W_2, W_3, W_4, x_{En}, G, G_R. Boundary of optimization is inequality

$$5 < G \le 20, \;\; 4 < G_R < 15,$$
$$1 < x_{En} \le 10, \; 15 < x_{En} \le 20$$
$$10 < W_i \le 90, \; i = 1..4$$

Important to note, that in the course of numerical experiment dimensionless quantities used. Also, scale of each variable had to be separately.

Calculation results

The procedure for calculating the objective function is to solve systems of equations and computing optimization criterion CR.
We studied the influence of the parameters of the modified random search on the result.
Empirical parameters of the algorithm are:
M- quantity of trajectories of search. Parameter M is defendant on the size and structure of the set of optimization H.
a- parameter that determines the magnitude of the neighborhood node trajectory, that is the step length. Depends on the size of the set X, the number of trajectories m, may change during the search.
Ψ - Hyper-cone opening angle. Also depends on the number of trajectories, also depends from the value of a.
M_k - the number of realizations of a random vector in the vicinity of the site. Depends on the value of a.
l-number of search steps for a partial path, which should be carried out after the screening of unpromising paths. Depends on the size of the set X.
r_x, r_y-distance threshold for clipping algorithm. Depend on the size of the set X.
p-length algorithm, that is, the number of search steps, after passing which the algorithm terminates.
The number of random points in the vicinity of each site recorded m_k = 25.
Table 1 shows the dependence of the value of global maximum on the parameters of the modified random search. This is a summary table of all the results of the numerical experiment. The experiment was conducted as follows.

At first, the parameters were chosen based on the size of the set of optimization. The sets which give the best and the worst results of the global search were selected.
After that each value l, r_x, r_y, ψ varied in turn. Values A and M varied together with the ratio A/M.
Only the most "noticeable" (best or worst) results are included in this table.

Table 1

Dependence on the parameters of the search result

№	M	A	l	r_x	r_y	ψ	p	CR
1	40	70	2	15	40	60	80	432
2	50	70	3	15	50	45	80	546
3	70	50	3	18	70	40	80	560
4	60	50	4	18	85	40	90	597
5	50	70	2	15	40	45	90	446
6	60	70	3	15	50	45	90	539
7	50	50	3	18	70	40	90	535
8	50	50	4	18	85	30	90	555
9	50	70	4	15	50	45	80	576
10	50	70	3	15	50	40	80	602
11	30	40	3	18	70	40	90	568
12	30	50	4	18	85	30	100	501

At the first and last rows of the table are differenced from the remaining rows by ratio of A to M (greater than 1.5). Values of the search results in these lines are the worst. It is clear that the parameters A and M, defining uniform distribution at each step have the greatest impact on the search process. It seems likely, that distribution parameters depend on the measure defined on the space containing the set of optimization.
Results obtained by varying the parameter are noteworthy. This parameter specifies how often the search is required to compare the trajectory, check their "closeness". Worst results were obtained for $l = 2$, that is, the comparison is made in 2 steps. It can be assumed that the frequent "rejection" close trajectories increases the probability of "loss" of the global optimum.
It should also be noted that the changing of the opening angle of the cone does not entail a significant improvement or deterioration in the search result. Possible, if we had a priori information about the objective function, we could give specific advice about it.
In row 10 of the table there is a set of parameters, which is the maximum value of CR. In this series of calculations, it was a path leading to the global optimum consisting of 46 steps. In general, total amount of trajectories was 50 , 23 of them were dropped as "close", 17 were located outside of the set of the optimization, 4 were found in the neighborhood of the optimum. It is interesting, the trajectory with the parameters of the 4 row of the table1 determined a similar value is the global maximum of 49 nodes. Thus, significant increase of the number of initial trajectories does not affect the duration of the search, because increases the number of dropped "close" trajectories. In this case, from 70 of trajectories were discarded 41, although the comparison at every forth step was carried out. So, influence of parameters which define rule of clipping paths on a search result are not revealed. Search results may be impaired, if, obviously, to reduce greatly the total number of trajectories.

Conclusion

In conclusion, we note that the global optimization methods based on random search very effective in the right choice of distributions. This methodology can be very similar to that of genetic search using the procedures of "selection" and "mutation". In addition, the idea of "crossing" the most appropriate one can be used for the construction of the distribution of random points at each step. Composition of these approaches can provide some interesting results. Comparison of computational efficiency requires a detailed study.

References

Ananchenko, A.G., Kholodnov, V.A., Punin, A.Ye. 2004 Globalnaya optimizatsiya cascade reactorov (Global optimization of cascade reactors). *Izvestya visshikh uchebnih zavedenii*, 47, pp.25-29. (In Russian)

Kholodnov, V.A., Sirenek, V.A., 1997. Mathematische Modellierung des instationa-ren Auslaugens von Glas unter Beachtung der Verzogerung Diffusionstrome.(Mathematical modeling of instating acids leaching of glass with respect to the delay diffusion currents) *Wissenschaftliche Zeitschrift*, TH Leuna-Merseburg, 29,1987,№4, S. 470-478.

Lebedeva, M.Yu., Kholodnov ,V.A., 2004. optimizatsiya himiko-tekhnologicheskoy sistemi v usloviyah intervalnoy neopredelennsty.(Optimization of Chemic-tecnological system under conditions of interval uncertainly). *Izvestya visshikh uchebnih zavedenii*, 49, pp.120-125. (In Russian).

Neumaier, A., 2000. Global Optimization Techniques. 2000 Available at: http://www.mat.univie.ac.at/~neum/glopt/techniques.html

Zabinsky, Z.B., 1991 Stochactic adaptive search for global optimization. Available at: http://courses.washington.edu/inde510/516/ZabinskyBookCh1.pdf

Weise, T. 2009 The global optimization: theory and application Available at: http://www.it-weise.de/projects/book.pdf

The analysis of effectiveness of the shortest path for traveling salesman problems in 2D grid

Edvards Valbahs, Peter Grabusts

Faculty of Engineering, Rezekne Higher Education Institution, Rezekne, Atbrivosanas aleja 90, Latvia
err@inbox.lv, peter@ru.lv

Abstract: *In order to achieve the wide range of the robotic application it is necessary to provide iterative motions among points of the goals. For instance, in the industry mobile robots can replace any components between a storehouse and an assembly department. Ammunition replacement is widely used in military services. Working place is possible in ports, airports, waste site and etc. Mobile robots can be used for monitoring if it is necessary to observe control points in the secret place. The paper deals with path planning programme for mobile robots. The aim of the research paper is to analyse motion-planning algorithms that contain the design of modelling programme. The programme is needed as environment modelling to obtain the simulation data. The simulation data give the possibility to conduct the wide analyses for selected algorithm. Analysis means the simulation data interpretation and comparison with other data obtained using the motion-planning. The results of the careful analysis were considered for optimal path planning algorithms. The experimental evidence was proposed to demonstrate the effectiveness of the algorithm for steady covered space. The results described in this work can be extended in a number of directions, and applied to other algorithms.*

Keywords: robotic, Simulated Annealing, path planning.

Introduction

The article is connected to the travelling salesman problem (TSP), but with some exceptions and conditions. In the case when the TSP is envisaged the following approximate path planning algorithms are used (Applegate, 2007; Cook, 2011; Davendra, 2010):

- The closest neighbour algorithm;
- Simulated Annealing (SA);
- Genetic Algorithm (GA);
- Ant colony optimization.

The closest neighbour approach is the simplest and straightforward TSP one (Johnson, 1995). The way to this approach to always visit the closest city. The polynomial complexity of the approach is $O(n^2)$. The algorithm is relatively simple:

1. Choose a random city.
2. Find out the nearest city unvisited and visit it.
3. Are there any unvisited cities left? If yes, repeat step 2.
4. Return to the first city.

SA is successfully used and adapted to get an approximate solutions for the TSP (Johnson, 1995). SA is basically a randomized local search algorithm similar to Tabu Search but do not allow path exchange that deteriorates the solution. The polynomial complexity of the approach is $O(n^2)$ accordingly.

```
Input: ProblemSize, iterations_max, temp_max
Output: S_best
S_current ← CreateInitialSolution(ProblemSize)
S_best ← S_current
for i = 1 to iterations_max do
  S_i ← CreateNeighborSolution(S_current)
  temp_curr ← CalculateTemperature(i, temp_max)
  if Cost(S_i) ≤ Cost(S_current) then
    S_current ← S_i
    if Cost(S_i) ≤ Cost(S_best) then
      S_best ← S_i
    end
  else if Exp((Cost(S_current)-Cost(S_i))/temp_curr) > Rand() then
    S_current ← S_i
  end
end
return S_best
```

Fig. 1. **Pseudocode for SA**.

The SA method (Otten, 1989; Kirkpatrick, 1983) is widely used in applied science (Fig. 1). The well-known traveling salesman problem has effectively solved by means of this method. For instance, the arrangement of many circuit elements on a silicon substrate is considerably improved to reduce interference among the elements (Kirkpatrick, 1984; Vecchi, 1983).

GA conducts in a way similar to the nature (Cook, 2011). A basic GA statrs working with a randomly generated population of potential solution. The candidates are then mated to produce offspring and only some of them go through a mutating process. Each candidate has an optimal value demonstrating us how go it is. Choosing the most optimal candidates for mating and mutation the overall optimality of the candidate solutions increases. Using GA to the TSP involves implementing a crossover procedure, a measure of optimality and mutation as well. Optimality of the solution is a length of the solution.

Ant colony optimization is the algorithm that is inspired by the nature (Dorigo, 1996). It is based on ant colony moving behaviour. Good results can be achieved by means of the algorithm but not for complex problems.

We managed to use SA algorithm rather successfully in our previous work (Valbahs, 2012) taking into account the specific side of this work (it will be discussed in detail further). Therefore, it is necessary to discuss some principles of SA realization in detail. In order to calculate the total path it is necessary to know the shortest route among all the cities. As we do not know the distance, we must apply one of the algorithms to define the shortest route among all the cities. It is Dijkstra's algorithm (Dijkstra, 1959) that gives the possibility to get the shortest path tree. The polynomial complexity of the Dijkstra's algorithm is $O(n^2)$.

Goals

The aim of the research paper is to analyze motion-planning algorithms that contain the design of modelling programme. The programme is needed as environment modelling to obtain the simulation data. The simulation data give the possibility to conduct the wide analyses for selected algorithm. Analysis means the simulation data interpretation and comparison with other data obtained using the motion-planning.

The use in practice and the necessity of it is greatly connected to optimal algorithm and methodological work out for autonomous agents that move in the space and are able to plan route on their own (Siegwart, 2011; Batavia, 2000; Biswas, 2002; Choset, 2000; Ashkenazi, 2000; Buhmann, 1995). One of such agent-samples exiting in our everyday life is autonomous vacuum cleaner. Autonomous vacuum cleaners do not usually use the motion-planning algorithm. They are based on some simple algorithms, for example cleaning in a spiral, crossing the premises avoiding the walls and their moving is casual after touching the walls. The philosophy of this design was offered by the scientists of Massachusetts Institute of Technology. Agents must behave as insects having primitive controlling devices in accordance to the environment. As a result, though an autonomous vacuum cleaner is very effective in cleaning premises, it is required much more time as compared with work made by a human. There is a drawback, the autonomous vacuum cleans some spaces many times but other spaces only once. The use of motion-planning algorithms can raise the effectiveness of an autonomous vacuum cleaner.

Assumptions

In order to fulfill the aim of the research paper the following conditions are introduced for:

- premises where an object moves;
- robot (or object) moves around the premises;
- path the robot moves on in the premises.

The premises are presented as two-dimensional plane. The plane of premises is equally divided into the cells. The cell dimensions are equal to agent dimension that moves in the premises. The space can be represented as a graph with two kinds of edges (Fig. 2). Horizontal and vertical edges are marked with unbroken lines they are of similar length, but others are longer and marked with dash lines. It is linked with agent movement possibilities.

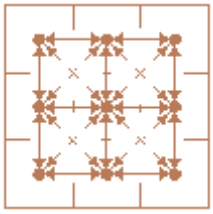


Fig. 2. **The example of the graph and 3 x 3 space**.

The object moves only one cell forward and back i.e. during one motion the object can move to the one cell from empty eight ones (eight cells around one cell) paying attention to that cell is not occupied by the obstacle but if it is occupied, the graph will not have the relevant vertex (Fig. 3).

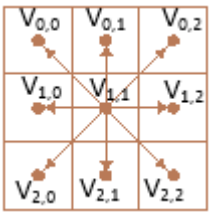


Fig. 3. **The example of agent's motion (where $v_{i,j}$ is relevant vertex)**.

As opposed to classical TSP we take a number of additional rules and it means that the agent can cross the one the same cell several times in succession (it must cross any cell one time obligatory). Thus, the agent's initial vertex does not coincide with its final vertex of total route.

In this research paper both algorithms were compared practically using and combining different placement of obstacles in the unchangeable two-dimensional space. All the results were obtained on one and the same computer (2.66 GHz processor and 2GB RAM), operating systems (Ubuntu 12.04.1 LST Linux were used). The following information was collected about:

- the number of covering for each cell;
- the time which was necessary for both algorithms to plan the route.

Results

Taking into account the fact that the distance among all the vertexes (cities) are unknown in the beginning, it is necessary to define the shortest paths among those vertexes mentioned above. Dijkstra's algorithm can be used but increasing the measures of the premises, the time is proportionally increases accordingly that is necessary for evaluating path tree. Therefore, it is needed to simplify the calculation of the shortest path, which is possible, provided the peculiarities and nuances of the task are taken into consideration. In addition, the empty premises should be observed. If all the mentioned above remains valid, the simple algorithm can be worked out to define the shortest paths.

Let us consider the agent's general moving paths. If there are no vertexes between the current initial and goal vertexes, the agent can move only to eight possible positions (cells) depending on goal vertex (Fig. 3). Admitting that first vertex index *i* defines the vertical position and the second vertex j defines the horizontal position we can draw a line either horizontally or vertically. And one of the vertexes will have the index with common value (Fig. 4).

Fig. 4. **The example of agent moving horizontally (where *i* index value is common for both vertexes)**.

Another situation can be seen if the current initial and goal vertexes are neither on the horizontal nor vertical lines (Fig. 5).

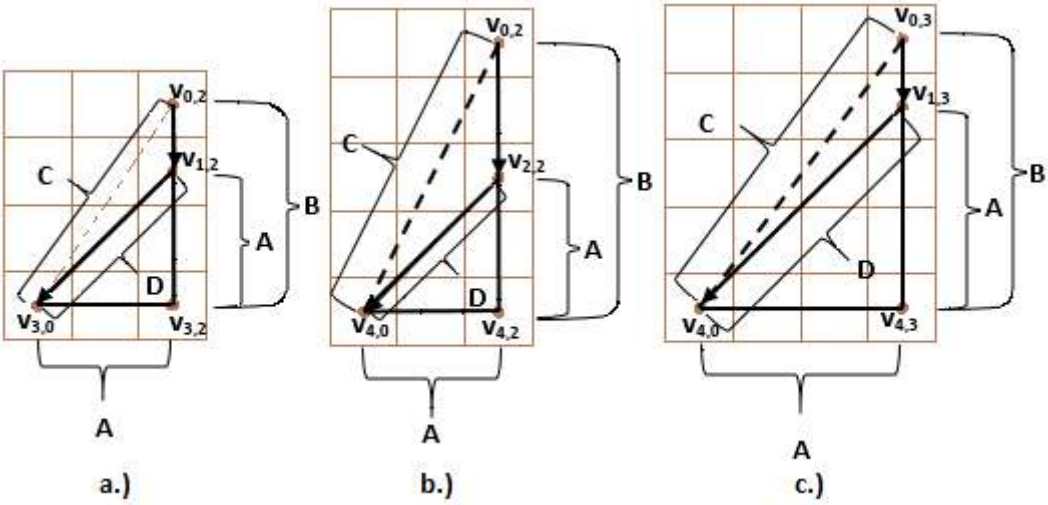


Fig. 5. **Three examples of agent moving (where *A*, *B* and *C* are sections among the vertexes): a.) agent moves from $v_{0,2}$ to $v_{3,0}$ crossing $v_{1,2}$; b.) agent moves from $v_{0,2}$ to $v_{4,0}$ crossing $v_{2,2}$; c.) agent moves from $v_{0,3}$ to $v_{4,0}$ crossing $v_{1,3}$)**.

All cases of Fig. 5 have common characteristics that unites them. The shortest path from initial vertex to goal vertex is section *C* but for the agent this path is unavailable because of current task conditions and peculiarities. These cases can be described by the right-angled triangle where *C* is a side of the triangle. In addition, side *B* is longer than side *A*.
One of the shortest paths among the relevant (corresponding) vertexes:

- the agent moves along the longest side B of the right-angled triangle until the gap between the covered path and side B is equal to side A;
- if gap between the covered path and side B is equal to side A, then the agent moves along the angle allowed (along the section D) to the goal vertex (let us mark that this action corresponds to the case when side B is equal to side A i.e. the right-angled triangle is the isosceles triangle, too (see Fig. 5 a.) in case initial vertex is $v_{1,2}$, 4 b.) in case initial vertex is $v_{2,2}$ and 5 c.) in case initial vertex is $v_{1,3}$)).

We can follow that the path is longer than optimal side C. And it can be calculated by the use of following formulae: $L = B-A+2^{0.5}*A$, where L is the length of the path from initial vertex to goal vertex. By turn, C can be calculated from $C = (A^2+B^2)^{0.5}$. It is possible to calculate how match percent L is longer than C (if L is equal to 100 %), then the final result is equal to P=((L-C)*100)/L. Our goal premises are 100 x 100 cells. The value of P is reflected with contour line for the given premises depending on A and B (Fig. 6).

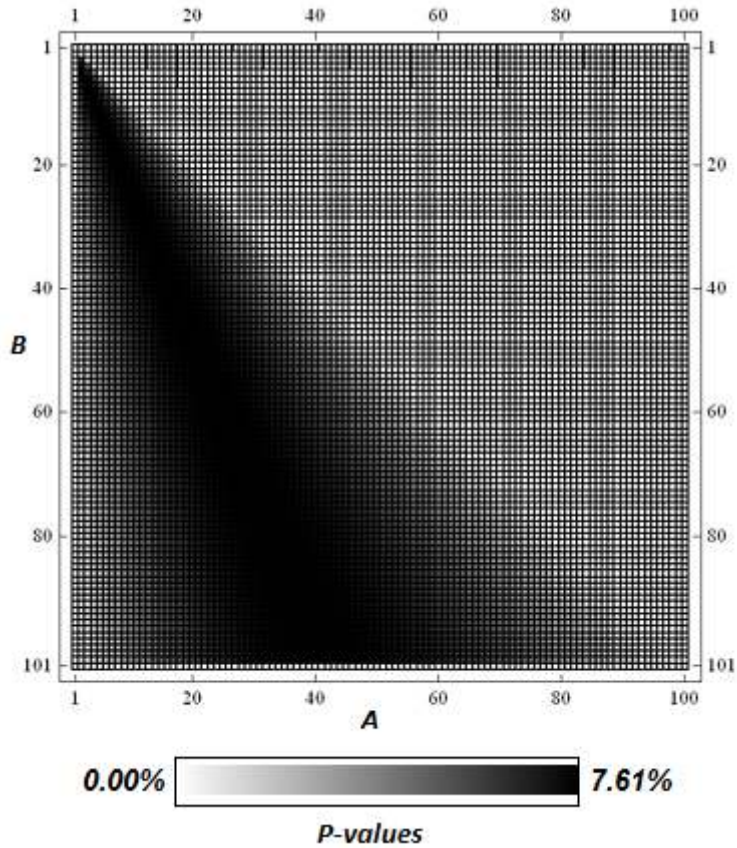


Fig. 6. ***P* value depending on *B* and *A*, if *A* > *1* and *B* > *A***.

It is possible to calculate maximum *P* value for 100 x 100 cells big premises (Fig. 6) that is equal to 7.61 %. The method/algorithm mentioned was applied instead of Dijkstra's algorithm to calculate total path or covering of 100 x 100 cells big premises and it is obstacles free (see Fig. 7).

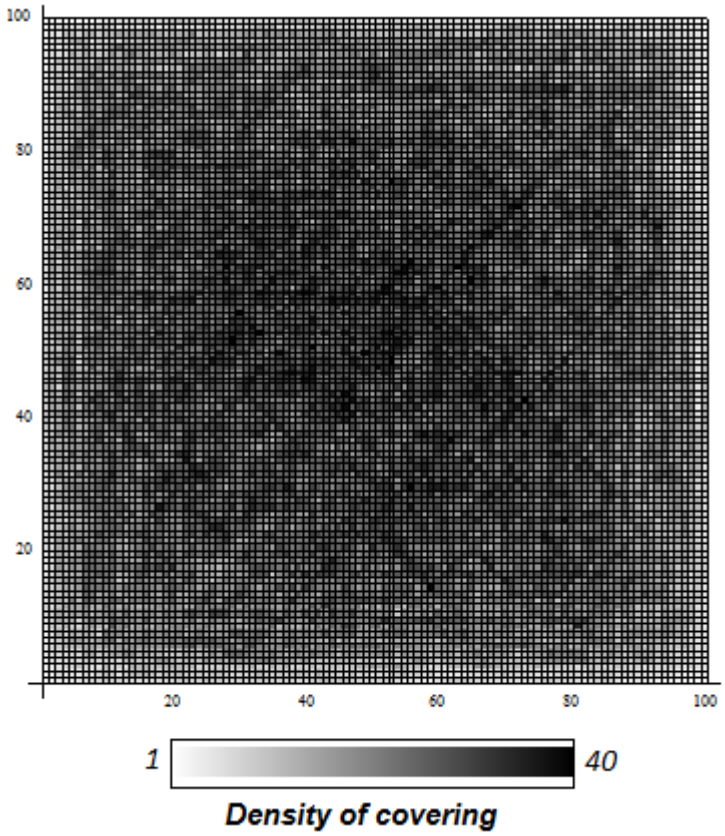


Fig. 7. **The density of covering for 100 x 100 cells big premises (it is obstacles free)**.

Density of covering changes from 1 up to 40 (there are the cells which were covered only once and there are the cells covered maximum 40 times). Totally, the agent performed 192666 steps in order to cross each cell of the premises.

Conclusions

It can be concluded that the algorithm offered is rather simple and it replaced Dijkstra's algorithm effectively according to the task. The algorithm allows decreasing the time of calculation, which is necessary to define the shortest route among graph vertexes.

The shortest path can be defined in a simple way (even in such cases mentioned in Fig. 5), provided that it is necessary to know the gap between the indexes of initial and goal vertexes. For instance, if initial vertex is $v_{i1,j1}$ and goal vertex is $v_{i2,j2}$, the first gap is $\Delta_1=|i_1-i_2|$ and the second gap is $\Delta_2=|j_1-j_2|$. As to the next step, it is needed to calculate the biggest gap between both the gaps. The shortest path is equal to the biggest gap. For instance, Fig. 5 a.) reflects the shortest path which occupies 4 cells, but in other cases it is 5 cells big.

It must be marked that total path can be a bit longer it is connected to the specific task which was envisaged in the chapters “Assumptions” and “Results” in detail. The worst case can be evaluated theoretically for the premises of 100 x 100 cells. If we take into consideration that the total route will consist of path pieces, which are longer than 7.61 % in comparison with *C* value (see Fig. 6), the total path will be longer than optimal 7.61 % (actually, it is the worst maximal variant. We must pay attention to the fact that SA provides only approximate solution).

The algorithm can be successfully used e.g. in autonomous public transport restricted by means of rules, technical requirements in autonomous robots and military equipment. In addition, the algorithm can be used in various computer games where a path planning is done in dynamic environment.

It is possible to conclude that the algorithm offered can be used in the different application areas not only for path planning of a robot.

References

Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J., 2007. The Traveling Salesman Problem, *Princeton University Press*, Princeton, USA.

Ashkenazi, V., Park, D., Dumville, M., 2000. Robot Positioning and the Global Navigation Satellite System, *Industrial Robots: An International Journal*, 27(6), pp. 419-426.

Batavia, P.H. and Nourbakhsh, I., 2000. Path planning for the Cye personal robot, *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS).

Biswas, R., Limketaki, B., Sanner, S., Thrun, S., 2002. Towards Object Mapping in Dynamic Environments with Mobile Robots, In:Proceedings of the Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland.

Buhmann, J., Burgard, W., Cremers, A.B., Fox, D., Hofmann, T., Schneider, F., Strikos, J., Thrun, S., 1995. The Mobile Robot Rhino, *AI Magazine*, 16(1).

Choset, H., 2000. Coverage of Known Spaces: The Boustrophedon Cellular Decomposition in Autonomous Robots, 9:247-253, *Kluwer*.

Cook, W.J., 2011. In Pursuit of the Traveling Salesman, *Princeton University Press*, Princeton, USA.

Davendra, D., 2010. Traveling Salesman Problem, Theory and Applications, In:Tech, Rijeka, Croatia.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs, *Numerische Mathematik*, v. 1, p. 269-271.

Dorigo, M. and Gambardella, L.M., 1996. Ant Colonies for the Traveling Salesman Problem, *University Libre de Bruxelles*, Belgium.

Johnson, D.S. and McGeoch, L.A., 1995. The Traveling Salesman Problem: A Case Study in Local Optimization, November 20.

Kirkpatrick, S., Gelatt C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing, *Science*, 220, pp. 671-680.

Kirkpatrick, S. 1984. Optimization by Simulated Annealing: Quantitative Studies, *Journal of Statistical Physics*, 34, pp. 975-986.

Otten, R.H.J.M. and Ginneken, L.P.P.P., 1989. The Annealing Algorithm. *Kluwer Academic Publishers*.

Siegwart, R., Nourbakhsh, I.R., Scaramuzza, D., 2011. Introduction to Autonomous Mobile Robots, *A Bradford Book The MIT Press Cambridge*, Massachusetts London, England.

Valbahs, E. and Grabusts, P., 2012. Motion Planning of an Autonomous Robot in Closed Space with Obstacles, *Scientific Journal of RTU*. 5. series., Datorzinatne. - 15. vol., pp. 52-57.

Vecchi, M.P. and Kirkpatrick, S., 1983. Global Wiring by Simulated Annealing, *IEEE Transaction on Computer Aided Design*, CAD-2, pp. 215-222.

ALPHABETICAL LIST OF AUTHORS

AHILCENOKA Dace
Riga Technical University, Faculty of Computer Science and Information Technology
Address: Meza street 1/3, Riga, LV-1048, Latvia, e-mail: dace.ahilcenoka@rtu.lv

ALBERTS Maris
Institute of Mathematics and Computer Science of the University of Latvi
Address: Riga, Raiņa blvd. 29, LV-1459

ALEKSEJEVA Ludmila
Department of Modelling and Simulation, Institute of Information Technology, Riga Technical University
Address: Kalku str.1, LV-1658 Rīga, Latvia, e-mail: ludmila.aleksejeva_1@rtu.lv

ARHIPOVA Irina, Professor, Dr. sc. ing
Department of Control Systems, Faculty of Information Technologies, Latvia University of Agriculture
Address: Liela Str. 2, Jelgava, LV 3001, Latvia, e-mail: irina.arhipova@llu.lv

ASMUSS Julija
Institute of Telecommunications, Riga Technical University
Address: Azenes str. 12, LV-1048, Riga, Latvia, e-mail:julija.asmuss@rtu.lv

BAJOVS Andrejs
Riga Technical University, Faculty of Computer Science and Information Technology
Address: Meza street 1/3, Riga, LV-1048, Latvia, e-mail: andrejs.bajovs@rtu.lv

BARTUSEVICS Arturs
Faculty of Computer Since and Information Technology, RTU
Address: Kaļķu iela 1, Rīga LV-1658, Latvija, arturik16@inbox.lv

BONDARENKO Andrejs
Institute of Information Technology Riga Technical University
Address: Meza street 1/3, Riga, LV-1048, Latvia, e-mail: andrejs.bondarenko@gmail.com

BORISOV Arkady
Institute of Information Technology Riga Technical University, Latvia
Address: 1 Kalku Street, Riga, LV-1658, Latvia, e-mail: arkadijs.borisovs@cs.rtu.lv

BRIKMANE Liga
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459

BUMANIS Nikolajs
Latvia University of Agriculture, Faculty of Informational Technology
Address: Liela Str. 2, Jelgava, LV 3001, Latvia, e-mail: my.way@inbox.lv

BULINS Zigmunds
Institute of Applied Software, Riga Technical University
Address:Meza 1/3, Riga, Latvia, e-mai:Zigmund.Bulinsh@gmail.com

BULIPOPA Natalja
Faculty of Information Technologies, Latvia University of Agriculture, Latvia
Address: Liela Str. 2, Jelgava, LV 3001, Latvia, e-mail: cvetkova.natalja@gmail.com

BURCEVA Olga
Department of Modelling and Simulation, Riga Technical University
Address: 1 Kalku Street, Riga, LV-1658, Latvia, e-mail: olgaa.burceva@gmail.com

BUTANE Linda
Latvia University of Agriculture, Faculty of Information Technologies
Address: Liela Str. 2, Jelgava, LV 3001, Latvia, e-mail: butane.linda@gmail.com

CABS Kaspars
Riga Technical University
Address: Meza street 1/3, Riga, Latvia, LV-1048, e-mail: kaspars.cabs@gmail.com

CERNENOKS Juris
Faculty of Physics and Mathematics, University of Latvia, Riga
Address: Zellu street 8, Riga, LV – 1002, Latvia, e-mail: juris2x@inbox.lv

CEVERE Rudite, Asoc.prof., Dr.sc.comp
Department of Computer Systems, Faculty of Information Technologies, Latvia University of Agriculture,
Address: Liela Str. 2, Jelgava, LV 3001, Latvia, e-mail: rudite.cevere@exigenservices.com

CIBULIS Andrejs
Faculty of Physics and Mathematics, University of Latvia,
Address: Zellu street 8, Riga, LV – 1002, Latvia, e-mail: cibulis@lanet.lv

CIUCU Stefan
IT Department, "Nicolae Titulescu" University
Address: Calea Vacaresti 185, sector 4, Bucharest, Romania, e-mail: stefanciucu@univnt.ro

CIZOVS Jurijs
Department of Modelling and Simulation, Riga Technical University
Address:1 Mezha street, Building 4, Room 456, Riga, LV-1048, Latvia

CZAPLEWSKI Cezary
University of Gdansk
Addrress: ul. Sobieskiego 18, 80-952 Gdansk, Poland

DRAGOESCU Raluca
Statistics and Econometrics Department, The Bucharest University of Economic Studies
Address: 6, Piata Romana, 1st district, Bucharest, Romania, e-mail: dragoescuraluca@gmail.com

DUBURS Gunars
Latvian Institute of Organic Synthesis
Address: Aizkraukles str. 21, Riga LV1006, Latvia

DUKA Vita
[1] Latvia University of Agriculture, Liela str. 2, Jelgava, LV3001, Latvia
[2]Latvian Institute of Organic Synthesis, Aizkraukles str. 21, Riga LV1006, Latvia, e-mail:vita.duka@llu.lv

DZERVE Andris
Institute of Mathematics and Computer Science of the University of Latvia
Address:, Raiņa blvd. 29, Riga LV-1459, Latvia

FINASKINS DMITRIJS
Telecommunications institute, Riga Technical University
Address:Azenes street 12, LV-1048, Riga, Latvia, e-mail:dmitrijs.finaskins@rtu.lv

GAILUMS Aleksandrs
Faculty of Information Technologies Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: aleksandrs.gailums@llu.lv

GALININA Anna
Department of Modelling and Simulation, Riga Technical University
Address: 1 Kalku Street, Riga, LV-1658, Latvia, e-mail:asr-89@mail.ru

GASPAROVICA Madara
Department of Modelling and Simulation, Institute of Information Technology, Riga Technical University
Address: Kalku str.1, LV-1658 Rīga, Latvia, e-mail: madara.gasparovica@rtu.lv

GORSKIS Henrihs
Department of Modelling and Simulation, Riga Technical University
Address: 1 Mezha street, Building 4, Room 456, Riga, LV-1048, Latvia, e-mail: Henrihs.Gorskis@rtu.lv

GORBIKS Olegs
Faculty of Computer Science and Information Technology, Riga Technical University,Latvia
Address: 1 Mezha street, Building 4, Riga, LV-1048, latvia, e-mail:olegs.gorbiks@rtu.lv

GRABIS Janis
Tele2 AB, Riga Technical University
Address: 1 Mezha street, Building 4, Riga, LV-1048, Latvia, e-mail: grabis@rtu.lv

GRABUSTS Peter
Engineering Faculty, Rezekne Higher Educational Institution
Address: Atbrivoshanas alley 90, Rezekne, LV-4601, Latvia, e-mail:peter@ru.lv

GRINBERGS Ugis
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459, Latvia, e-mail: ugis.grinbergs@lumii.lv

GUSAROVS Konstantins
Faculty of Computer Science and Information Technology, Riga Technical University
Address: 1 Mezha street, Building 4, Riga, LV-1048, Latvia, e-mail:konstantins.gusarovs@rtu.lv

JEKABSONS Gints
Institute of Applied Computer Systems, Riga Technical University
Address: Meza street 1/3, Riga, Latvia, LV-1048

JEKABSONS Vilnis
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459

KALEJS Andris
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459

KAMINSKIENE Janina
Center of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University
Address: Universiteto 10, Kaunas, Lithuania

KASKATAYEVA Bakhytkul
Kazakh National Peadagogical University Abai
Address: 13 Dostyk ave., Almaty, 050010, Republic of Kazakhstan

KODORS Sergejs
Rezekne Higher Education Institution
Address: Brivibas street 39-9, Rezekne, LV-4601, Latvia, e-mail: sk_7@inbox.lv

KOPEIKA Evija
Faculty of Economics and Management, University of Latvia
Address: Aspazijas Boulevard 5 - 332, Riga, LV -1050, e-mail:evija.Kopeika@llu.lv

KOZACENKO Ludmila
Riga Technical University, Faculty of Computer Science and Information Technology
Address: Meza street 1/3, Riga, LV-1048, Latvia, e-mail: ludmila.kozacenko@rtu.lv

KRASOVSKIS Viktors
Faculty of Engineering, Rezekne University College of Applied Sciences, IT
Address: Atbrīvošanas aleja 115, Rēzekne, LV-4601, Latvia, e-mail: viktors.krasovskis@gmail.com

KREBS Valdis
Institute of Mathematics and Computer Science, University of Latvia
Address: Raina bulvaris 29, Riga, LV-1459, Latvia, e-mail: valdis.krebs@lumii.lv

KREISMANE Dzidra
Latvia University of Agriculture , Jelgava
Address: Liela str. 2, LV-3001, Latvia, e-mail: dzidra.kreismane@llu.lv

KVIESIS Armands
Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: armands.kviesis@inbox.lv

LAUKS Gunars
Institute of Telecommunications, Riga Technical University
Address: Azenes str. 12, LV-1048, Riga, Latvia, e-mail: gunars.lauks@rtu.lv

LEIKUMS Toms
Faculty of Information Technology, Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: Toms_Leikums@programmer.net

LEKTAUERS Arnis
Riga Technical University
Address:1 Kalku Street, LV-1658 Riga, Latvia, e-mail: arnis@itl.rtu.lv

LIEPINA Inta
Latvian Institute of Organic Synthesis
Address: Aizkraukles str. 21, Riga LV1006, Latvia, e-mail: *inta@osi.lv*

MACHOVIKOV Alexei
Department of Informatics and Computer Technologies, National University of Mineral Resources 'Mining', 2, 21st Line, St-Petersburg, Russia, 199106
Department of Software Development, Bradon Technologies Ltd., 5225 Orbitor Drive, Unit 22, Mississauga, Ontario, Canada, L4W 4Y8, e-mail:amachovikov@gmail.com

LIWO Adam
University of Gdansk
Address: ul. Sobieskiego 18, 80-952 Gdansk, Poland

MENSAH Peter
Department of Modeling and Simulation, Riga Technical University
Address:1 Kalku Street, LV-1658 Riga, Latvia, e-mail: p_mensah@hotmail.com

MERKURYEV Yuri
Riga Technical University
Address: 1 Kalku Street, LV-1658 Riga, Latvia , e-mail: merkur@itl.rtu.lv

MIKULOVS Igors
Tele2 AB, Riga Technical University
Address: 1 Kalku Street, LV-1658 Riga, Latvia, e-mail:igors.mikulovs@tele2.com

NAMIOT Dmitry
Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, e-mail:
Russia dnamiot@gmail.com

NAZARUKS Vladislavs
Department of Applied Computer Science, Riga Technical University
Address: Kalku str.1, LV-1658 Rīga, Latvia, e-mail: vladislavs.nazaruks@rtu.lv

NIKIFOROVA Oksana
Faculty of Computer Science and Information Technology, Riga Technical University
Address: 1 Kalku Street, LV-1658 Riga, Latvia, e-mail:oksana.nikiforova@rtu.lv

NIKOLAJEVS Jurijs
Institute of Applied Computer Systems, Riga Technical University
Address: Meza street 1/3, Riga, Latvia, LV-1048, e-mail: juryn@inbox.lv

OANCEA Bogdan
Finance and Accounting Department, "Nicolae Titulescu" University
Address: Calea Vacaresti 185, sector 4, Bucharest, Romania, e-mail: bogdanoancea@univnt.ro

ODZINA Ilona, Mg.sc.ing.
Biosystems Group, Faculty of Information Technology, Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e:mail:odzinai@ibox.lv

OPMANIS Rihards
Institute of Mathematics and Computer Science, University of Latvia
Address: Raina bulvaris 29, Riga, LV-1459, Latvia, e-mail: rihardso@latnet.lv

OPMANIS Rudolfs
Institute of Mathematics and Computer Science, University of Latvia, Latvia
Address: Raina bulvaris 29, Riga, LV-1459, Latvia, e-mail:rudolfs.opmanis@gmail.com

OSADCUKS Vitalijs
Faculty of Engineering, Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: vitalijs.osadcuks@llu.lv

PAULINS Nauris
Information system department, Latvia Univerisity of Agriculture, Latvia
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail:Nauris.paulins@llu.lv

PAURA Liga
Department of Control System, Faculty of Information Technologies, Latvia University of Agriculture
Address: Liela str.2, Jelgava, LV-3001, e-mail: Liga.paura@llu.lv

PEVNEVA Anna
National Mineral Resources University
Address: 199106 21 line-3, Saint-Petersburg, Russia, e-mail: pevnevaa@inbox.ru

POLAKA Inese
Institute of Information Technology, Riga Technical University
Address: Meza ¼ - 463, Riga, Latvia, LV-1048, Latvia, e-mail: Inese.Polaka@rtu.lv

POULOVA Petra
Faculty of Informatics and Management, University of Hradec Kralove
Address: Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic, e-mail: petra.poulova@uhk.cz

PURVINS Ojaras
Kaunas University of Technology
Address: Klaipedos 3, KTU, Panevežys, Lithuania, e-mail: opurvi@panf.ktu.lt

PROVOROVA Irina
Department of Computer Science, Riga Technical University (RTU)
Address: Meza ¼ - 463, Riga, LV-1048, Latvia

PROVOROVS Sergejs
Department of Computer Science, Riga Technical University (RTU)
Address: Meza ¼ - 463, Riga, Latvia, LV-1048, e-mail: psv@psv.lv

RIMKUVIENE Daiva
Center of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University,
Address: Universiteto 10, Kaunas, Lithuania, e-mail: Daiva.Rimkuviene@lzuu.lt

RUPSYS Petras
Institute of Forest Management and Wood Sciences, Aleksandras Stulginskis University, Kaunas, LT-53361, Lithuania
Centre of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University, LT-53361, Lithuania, e-mail:Petras.rupsys@asu.lt

SAMUSENKO Ivans
Faculty of Engineering, Rezekne University College of Applied Sciences, IT
Address: Atbrīvošanas aleja 115, Rēzekne, LV-4601, Latvia

SAPATS Martins
Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail:martins.sapats@llu.lv,

SIMONOVA Ivana
Faculty of Informatics and Management, University of Hradec Kralove
Address: Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic,e-mail: ivana .simonova@uhk.cz

SITIKOVS Vjaceslavs
Institute of Applied Software, Riga Technical University
Address:Meza 1/3, Riga, Latvia, e-mail: Vjaceslavs.Sitikovs@rtu.lv

SKINDERE Nadezda
Riga Technical University, Faculty of Computer Science and Information Technology
Address: Meza ¼ - 463, Riga, LV-1048, Latvia, e-mail: nadezda.skindere@rtu.lv

SNEPS-SNEPPE Manfred
Ventspils International Radio Astronomy Centre Ventspils University College, Ventspils, Latvia
e-mail:manfreds.sneps@gmail.com

SPROGE Sandra
Latvia University of Agriculture, Faculty of Information Technology
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: Sandra.Sproge@llu.lv;

SUSNIENE Dalia
Kaunas University of Technology
Address: Klaipedos 3, KTU, Panevežys, Lithuania, e-mail: dalia.susniene@ktu.lt

TEILANS Artis
Rezeknes Higher Education Institution,
Address: 115 Atbrivosanas aleja, LV-4600,Rezekne, Latvia, e-mail: artis.teilans@ru.lv

TIKUMA Baiba
Latvia University of Agriculture , Jelgava
Address: Liela str. 2, LV-3001, Latvia

UNGURS Dainis
Riga Technical University, Faculty of Computer Science and Information Technology
Address: Meza ¼ - 463, Riga, LV-1048, Latvia, e-mail: dainis.ungurs.@rtu.lv

UZGA-REBROVS Olegs
Faculty of Engineering, Rezekne University College of Applied Sciences, IT
Address: Atbrīvošanas aleja 115, Rēzekne, LV-4601, Latvia

VALBAHS Edvards
Faculty of Engineering, Rezekne Higher Education Institution
Address: Rezekne, Atbrivosanas aleja 90, Latvia, e-mail: err@inbox.

VESELIS Normunds
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459

VINTERE Anna
Department of Mathematics, Latvia University of Agriculture
Address: Liela street 2, Jelgava, Latvia, LV-3001, e-mail: Anna.Vintere@llu.lv

ZEIDMANE Anda, Dr.paed.
Faculty of Information Technologies, Latvia University of Agriculture
Address: Liela Str.2, Jelgava, LV 3001, Latvia, e-mail: anda.zeidmane@llu.lv

ZMANOVSKA Tatjana
Department of Modelling and Simulation, Riga Technical University
Address: 1 Mezha street, Building 4, Room 456, Riga, LV-1048, Latvia

ZOTOVS Viktors
Institute of Mathematics and Computer Science of the University of Latvia
Address: Riga, Raiņa blvd. 29, LV-1459

Applied Information and Communication Technologies / Proceedings of the 6-th International Scientific Conference, Jelgava, Latvia, April 25-26, 2013. –327 p.