

# **Third International Conference**

## **INFORMATION RESEARCH, APPLICATIONS AND EDUCATION**

**27-30 June 2005, Varna**



**itech**

### **PROCEEDINGS**

**FOI-COMMERCE**

**SOFIA, 2005**

Kr. Markov (Ed.)

Proceedings of the Third International Conference “Information Research, Applications and Education” i.TECH 2005, Varna, Bulgaria

Sofia, FOI-COMMERCE – 2005

ISBN: 954-16-0034-4

First edition

Printed in Bulgaria by Institute of Information Theories and Applications FOI ITHEA  
Sofia -1090, P.O. Box 775

e-mail: [foi@nlcv.net](mailto:foi@nlcv.net)

[www.foibg.com](http://www.foibg.com)

All rights reserved.

© 2005 Krassimir Markov - Editor

© 2005 Krassimira Ivanova – Technical editor

© 2005 Institute of Information Theories and Applications FOI ITHEA - Preprint processing

© 2005 FOI-COMMERCE - Publisher

© 2005 For all authors in the issue

® i.TECH is a trade mark of Krassimir Markov

ISBN 954-16-0034-4

C\o Jusautor, Sofia, 2005

---

## Preface

The International Conference “Information Research, Applications and Education” (i.TECH) is organized as a part of “ITA 2005 - Joint International Scientific Events on Information Theories and Applications”.

The main organizer of the ITA 2005 as well as the i.TECH 2005 is the **International Journal on Information Theories and Applications (IJ ITA)**.

The aim of the conference is to be one more possibility for contacts for IJ ITA authors. The usual practice of IJ ITA is to support several conferences at which the IJ ITA papers may be discussed before submitting them for referring and publishing in the journal. Because of this, such conferences usually are multilingual and bring together both papers of high quality and papers of young scientists, which need further processing and scientific support from senior researchers.

i.TECH 2005 was combined with a specialized International Workshop on Business Informatics (Bi'05). Bi'05 was devoted to discussion of current research, applications and education regarding the basic directions of business informatics.

I would like to express my thanks to all who support the i.TECH 2005 and especially to the Natural Computing Group (NCG) (<http://www.lpsi.eui.upm.es/nncg/>) of the Technical University of Madrid, which is led by Mr. Juan Castellanos. The group is one of the foundational groups of the European Molecular Computing Consortium (<http://openit.disco.unimib.it/emcc/welcome.html>). NCG is involved in natural computing activities researches including: Molecular Computing (DNA Computing and Membrane Computing), Artificial Neural Networks (new architectures and learning strategies, Chaos controlling by Artificial Neural Networks, etc), Artificial Intelligence, Evolutionary algorithms, etc. NCG is constituted by members of some departments of the Technical University of Madrid, having its site at the Department of Artificial Intelligence of the Faculty of Computer Science. NCG has participated in different national and international projects including INTAS and V Framework Program of European Union (MolCoNet).

Let me thank the Program Committee of the conference for referring the submitted papers. Special thanks to Mr. Viktor Gladun and Mr. Avram Eskenazi – Co-chairs of the Program Committee.

i.TECH 2005 Proceedings has been edited in the Institute of Information Theories and Applications FOI ITHEA and published by FOI COMMERCE Co.

The i.TECH 2005 Conference found the best technical support in the work of the Technical editor Ms. Krassimira Ivanova and Organizing secretary Mr. Ilia Mitov.

To all participants of i.TECH 2005 I wish fruitful contacts during the conference days and efficient work for preparing the high quality papers to be published in the International Journal on Information Theories and Applications.

Varna, June 2005

Krassimir Markov

IJ ITA Editor in Chief

**i.TECH 2005 has been organized by:**

International Journal "Information Theories and Applications"  
 Institute of Information Theories and Applications FOI ITHEA (Bulgaria)  
 V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine  
 Institute of Mathematics and Informatics, BAS (Bulgaria)  
 Association of Developers and Users of Intelligent Systems(Ukraine)  
 National Academy of Sciences of Ukraine  
 Natural Computing Group of the Technical University of Madrid (Spain)  
 IBM Research Division  
 Varna Free University "Chernorizets Hrabar" (Bulgaria)  
 New Technik Publishing Ltd. (Bulgaria)

**Program Committee:**

Victor Gladun (Ukraine) - Chairman  
 Avram Eskenazi (Bulgaria) – co-chair  
 Krassimir Markov (Bulgaria) -co-chair  
 Jennifer Trelewicz ( USA ) – Bi'05 chair  
 Stoyan Poryazov ( Bulgaria ) – Bi'05 co-chair

Adil Timofeev (Russia)	Jury Zaichenko (Ukraine)	Neonila Vashchenko (Ukraine)
Alexander Gerov (Bulgaria)	Koen Vanhoof (Belgium)	Nikolay Lutov (Bulgaria)
Alexander Kuzemin (Ukraine)	Konstantin Gaindrik (Moldova)	Nikolay Zagoruiko (Russia)
Alexander Palagin (Ukraine)	Krassimir Manev (Bulgaria)	Peter Stanchev (USA)
Alexey Voloshin ( Ukraine )	Krassimira Ivanova (Bulgaria)	Petia Asenova (Bulgaria)
Alfredo Milani (Italy)	Laura Ciocoiu (Romania)	Plamen Mateev (Bulgaria)
Anna Kantcheva (Bulgaria)	Levon Aslanyan (Armenia)	Radoslav Pavlov (Bulgaria)
Arkady Zakrevskij (Belarus)	Luis Fernando de Mingo López (Spain)	Rumiana Kirkova (Bulgaria)
Frank Brown (USA)	Maria Bruseva (Bulgaria)	Stanimir Stoyanov (Bulgaria)
Hans Joachim Nern (Germany)	Maria Kasheva (Bulgaria)	Stefan Dodunekov (Bulgaria)
Hristina Daskalova ( Bulgaria )	Maria Nisheva (Bulgaria)	Tatiana Atanasova (Bulgaria)
Iliia Mitov (Bulgaria)	Martin Mintchev (Canada)	Tsvetanka Kovacheva (Bulgaria)
Irina Jeliazkova (Bulgaria)	Milena Dobreva (Bulgaria)	Valery Koval (Ukraine)
Ivan Popchev (Bulgaria)	Natalia Ivanova (Rusia)	Vitaliy Lozovskiy (Ukraine)
Juan Penuella Castellanos (Spain)	Nelly Maneva (Bulgaria)	Vladimir Ryazanov (Russia)

Technical editor: **Krassimira Ivanova**  
 Organizing secretary: **Iliia Mitov**

**The main topics of:****i.TECH 2005**

Applied program systems  
 Education informatics  
 Extreme programming  
 Hyper technologies  
 Information modelling  
 Information systems  
 Multimedia systems  
 Performance evaluation of computer and telecommunication systems  
 Quality of the programs  
 Software engineering  
 Statistical systems

**Bi 2005**

Theoretical foundations of business Informatics: conceptions, languages, modelling technologies, metrics and measurements, information visualization

Tools and applications: tools for business information modelling, applied systems for business information service, business information access methods and collaboration tools, business information security and reliability, new customer engagement models enabled by business informatics tools

Business informatics education (BIE): BIE conceptions, including relation to other/traditional fields of study; computer aided BIE, methodology of BIE

**Official languages of the conference are Bulgarian, Russian and English.**

**General sponsor of the i.TECH 2005 is FOI BULGARIA ( [www.foibg.com](http://www.foibg.com) ).**

## TABLE OF CONTENTS

### Data Models and Processing

Data Flow Analysis and the Linear Programming Model <i>Levon Aslanyan</i> .....	7
The Boundary Descriptors of the $n$ -dimensional Unit Cube Subset Partitioning <i>Hasmik Sahakyan, Levon Aslanyan</i> .....	12
Update-Retrieve Problem: Data Structures and Complexity <i>Jose Joaquin Erviti, Adriana Toni</i> .....	18
The Distributed System of Databases on Properties of Inorganic Substances and Materials <i>Nadezhda N.Kiselyova, Victor A.Dudarev, Ilya V.Prokoshev, Valentin V.Khorbenko, Andrey V.Stolyarenko, Dmitriy P.Murat, Victor S.Zemskov</i> .....	22
Software Development for Distributed System of Russian Databases on Electronics Materials <i>Valery Komysenko, Victor Dudarev</i> .....	27
Analyzing the Data in OLAP Data Cubes <i>Galina Bogdanova, Tsvetanka Georgieva</i> .....	33

### Knowledge Engineering

Classification of Biomedical Signals using the Dynamics of the False Nearest Neighbours (DFNN) Algorithm <i>Charles Newton Price, Renato J. de Sobral Cintra, David T. Westwick, Martin P. Mintchev</i> .....	39
Tracking Sensors Bottleneck Problem Solution using Biological Models of Attention <i>Alexander Fish, Orly Yadid-Pecht</i> .....	51
Neural Control Model of Chaotic Dynamic Systems <i>Cristina Hernández de la Sota, Juan Castellanos Peñuela, Rafael Gonzalo Molina, Valentín Palencia Alejandro</i> .....	58
Symbolic and Numeric Connectionist Models Solving NP-Problems <i>Luis Fernando de Mingo López, Francisco Gisbert</i> .....	66
From Textual to Computational Information Modelling <i>Jesús Cardeñosa, Carolina Gallardo, Eugenio Santos</i> .....	71
A New Approach for Eliminating the Spurious States in Recurrent Neural Networks <i>Víctor Giménez-Martínez, Carmen Torres, José Joaquín Erviti Anaut, Mercedes Perez-Castellanos</i> .....	79
On Contradiction Degrees between Two Fuzzy Sets, Revisited <i>Carmen Torres, Elena Castiñeira, Susana Cubillo, Victoria Zarzosa</i> .....	87
An Approach to Collaborative Filtering by ARTMAP Neural Networks <i>Anatoli Nachev</i> .....	95
Synthesis Methods of Multiple-valued Structures of Language Systems <i>Mikhail Bondarenko, Grigorij Chetverikov, Alexandr Karpukhin, Svetlana Roshka, Zhanna Deyneko</i> .....	102
Architecture and Principles of Control for Multi-Agent Telecommunicational Systems of New Generation <i>Adil V.Timofeev</i> .....	108

### Software Engineering

Applying Hierarchical MVC Architecture to High Interactive Web Applications <i>Micael Gallego-Carrillo, Iván García-Alcaide, Soto Montalvo-Herranz</i> .....	110
A Sensitive Metric of Class Cohesion <i>Luis Fernández, Rosalía Peña</i> .....	115
RKHS-Methods at Series Summation for Software Implementation <i>Svetlana Chumachenko, Ludmila Kirichenko</i> .....	124
Examination of Archived Objects' Size Influence on the Information Security when Compression Methods are Applied <i>Dimitrina Polimirova-Nickolova, Eugene Nickolov</i> .....	130

Визуализация на алгоритми и структури от данни <i>Ивайло Петков, Сергей Георгиев</i> .....	135
Нов подход към конструирането на блок-схемни езици <i>Стоян Порязов</i> .....	141
Симулация на някои основни грешки при измерване на оптични параметри чрез прилагане на обратна задача в оптиката и фазово-стъпков метод <i>Георги Стоилов</i> .....	143
<b>Informatics in Education and Cultural Heritage</b>	
Digitisation of Cultural Heritage: between EU Priorities and Bulgarian Realities <i>Milena Dobрева, Nikola Ikononov</i> .....	149
Programming Paradigms in Computer Science Education <i>Elena I. Bolshakova</i> .....	155
Tutoring Methods in Distance Course «Web-design Technologies» <i>Valeriy Bykov, Yuriy Zhook, Ganna Molodykh</i> .....	160
Учебная модель компьютера как база для изучения информатики <i>Евгений А. Еремин</i> .....	165
Historical Informatics in Practical Application: Virtual Museum of the Khazar State History <i>Boris Elkin, Alexander Kuzemin, Alexander Koshchy, Alexander Elkin</i> .....	169
Distributed Information Measurement System for Support of Research and Education in Optical Spectroscopy <i>Sergey Kiprushkin, Nikolay Korolev, Sergey Kurskov, Natalia Nosovich</i> .....	171
<b>International Workshop “Business Informatics”</b>	
Levels of Business Structures Representation <i>Katalina Grigорова, Plamenka Hristova, Galina Atanasova, Jennifer Q. Trelewicz</i> .....	181
Decision Support System for Investment Preference Evaluation under Conditions of Incomplete Information <i>Ivan Popchev, Irina Radeva</i> .....	189
Автоматизация отбора структурированной информации для подготовки управленческих решений <i>Андрей Д. Данилов</i> .....	190
Viable Model of the Enterprise – a Cybernetic Approach for Implementing the Information Technologies in Management <i>Todorka Kovacheva</i> .....	200
Analysis of Movement of Financial Flows of Economical Agents as the Basis for Designing the System of Economical Security (General Conception) <i>Alexander Kuzemin, Vycheslav Liashenko, Elena Bulavina, Asanbek Torojev</i> .....	204
Using ORG-Master for Knowledge Based Organizational Change <i>Dmitry Kudryavtsev, Lev Grigoriev, Valentina Kislova, Alexey Zablotsky</i> .....	210
<b>INFRAWEB Project</b>	
Semantic Web Service Development on the Base of Knowledge Management Layer - INFRAWEB Approach <i>Joachim Nern, Tatiana Atanasova, Gennady Agre, András Micsik, László Kovács, Janne Saarela, Timo Westkaemper</i> .....	217
Adjusting WSMO API Reference Implementation to Support More Reliable Entity Persistence <i>Ivo Marinchev</i> .....	223
INFRAWEB Capability Editor – A Graphical Ontology-Driven Tool for Creating Capabilities of Semantic Web Services <i>Gennady Agre, Peter Kormushev, Ivan Dilov</i> .....	228

---

---

## Data Models and Processing

---

---

### DATA FLOW ANALYSIS AND THE LINEAR PROGRAMMING MODEL<sup>1</sup>

Levon Aslanyan

**Abstract:** *The general discussion of the data flow algorithmic models, and the linear programming problem with the varying by data flow criterion function coefficients are presented. The general problem is widely known in different names - data streams, incremental and online algorithms, etc. The more studied algorithmic models include mathematical statistics and clustering, histograms and wavelets, sorting, set cover, and others. Linear programming model is an addition to this list. Large theoretical knowledge exists in this as the simplex algorithm and as interior point methods but the flow analysis requires another interpretation of optimal plans and plan transition with variate coefficients. An approximate model is devised which predicts the boundary stability point for the current optimal plan. This is valuable preparatory information of applications, moreover when a parallel computational facility is supposed.*

**Keywords:** *data flow algorithm, linear programming, approximation*

---

#### 1. Introduction

---

Data flow is a concept, traditionally appearing in the sensor based monitoring systems. Advanced global networks brought a number of novel applied disciplines intensively dealing with data flows. The network monitoring itself and optimal management of telecommunication systems, search engines with consequent data analysis, the network measuring instruments and network monitoring for security, etc. are the novel examples of data flow models. These deal with continuous data flows and unusual, non-finite and non-stored data set. In this case, the queries (the data analysis requests) are long-term and continuous processes in contrast to usual one-time queries. The traditional databases and data processing algorithms are poorly adjusted for the hard and continuous queries in data flows. This generates the necessity of new studies for serving continuous, multilayered, depending on time and subjected to indefinite behaviour of data flows [MM 2003]. Concerning the mentioned problem area, systems and algorithms are devised for different needs: real time systems, automation control systems, modelling processes, etc., but they are episodes in point of view of the formulated general problem. Traditional trade offs of such systems include one-pass and multi-pass algorithms, deterministic and randomized algorithms, and exact and approximate algorithms. Off-line algorithms solve a problem with full knowledge of the complete problem data. Online algorithms construct partial solutions with partial knowledge of the problem data, and update their solutions every time some new information is provided. In other words, they must handle a sequence of closely related and interleaved sub-problems, satisfying each sub-problem without knowledge of the future sub-problems. Standard examples of online problems include scheduling the motion of elevators, finding routes in networks and allocating cache memory. The usual way of measuring the quality of an online algorithm is to compare it to the optimal solution of the corresponding off-line problem where all information is available at the beginning. An online algorithm that always delivers results that are only a constant factor away from the corresponding optimum off-line solution, is called a competitive algorithm.

---

<sup>1</sup> The research is supported partly by INTAS: 04-77-7173 project, <http://www.intas.be>

The “incremental update” algorithmic model of data analysis [AJ 2001] modifies the solution of a problem that has been changed, rather than re-solving the entire problem. For example, partial change of conditions of a time-table problem must be force to only partial reconstruction of the table. It is obvious that it is possible to construct a theoretical problem, where any particular change brings to the full reconstruction of the problem. It is also clear that there are numerous problems, which are not so critical to the local transformations. It is an option to try to solve the given data flow problem by the mentioned incremental algorithms, moreover, in the specific conditions it is the only possible way for solving the problem including the data flows analysis.

Measuring the “variability”, “sortedness” and similar properties of data streams could be useful in some applications; for example, in determining the choice of a compression or sort algorithm for the underlying data streams. [MM 2003] have studied the bit level changes in video sequences and [AJ 2001] - the problem of estimating the number of inversions (the key element of the Shell type sorting algorithms) in a permutation of length  $n$  to within a factor  $1 \pm \varepsilon$ , where the permutation is presented in a data stream model. [MM 2003] proves the decreasing bit level changes of image pixels in video sequences and in [AJ 2001] - an algorithm obtained requiring the space  $O(\log n \log \log n)$ .

Sketching tools are usual for many data oriented applications. These include approximations of statistical parameters, histograms, wavelets, and other similar general descriptors. The simplest calculations for data streams serve the base statistical means like the averages and variations [AC 2003]. Other data flow descriptors also appear in publications: frequency moments [AM 1996], histograms [GG 2002], etc.

The paper below discusses an important applied model for the flow environments. We consider the linear programming mathematical problem, parameters of which are formed by data flows. In a moment it is assumed that the optimal plan is found and the coordinates of the target function are variable by the flow. In this case, there is an emerging question: which mechanisms are well suited to follow the coefficients variations in creating the configuration of the next resulting optimal plan. It is clear that the small changes of coefficients lead to simple changes of the current optimal plan, probably not requiring the total analysis of the problem by the complete flow information.

---

## 2. Linear Programming in Data Flows

---

Let's consider the linear programming problem in its canonical form:

$$\begin{aligned} \min c'x \\ Ax = b, \quad x \geq 0, \end{aligned}$$

where  $c \in R^n$ ,  $b \in R^m$ ,  $A$  is an  $m \times n$  full rank real matrix, and  $m < n$ . Without a loss of generality we may also suppose that  $b_i \geq 0$ ,  $i = \overline{1, m}$ . Particular examples of linear programming problem are given through the definition of coefficient values:  $a_{ij}, c_j, b_i$ , for  $i = \overline{1, m}; j = \overline{1, n}$ . Let's imagine a specific situation arising from application that the mentioned coefficients are changing in time. Such problems appear, for example, in data flow analysis.

Let us consider a data flow  $B(t, n)$ , which is finite but a very large-sized sequence of values  $b_1, b_2, \dots, b_n$ , where  $b_t, t = \overline{1, n}$  are certain structures. The data flows processing algorithms may use comparatively small storages than the input size. A limitation window is given in certain cases for processing separate data fragments. The time-dependent values of parameters, forming the applied problem model are formed as a result of algorithmic analysis. Unlike other natural and similar definitions, the variation of parameters is unpredictable here, as it has not probabilistic distribution and is not described by one or another property. Instead, it is considered that the variation takes place very slowly, because of the accumulation concept. In its turn, the applied problem demands to have ready answers to the certain questions in each time stamp.

There are two strategies: (1) solving a problem for each stage by all actual information which is practically impossible because of the large data sizes; and (2) structuring hereditary systems when the new data analysis is relatively easily integrated with the results of the previous data analysis.



We are going to consider the linear programming model in the mentioned conditions. The arbitrary variation of the coefficients is not allowed, instead, slow variations are considered so that the variation is fully monitored and it changes the solutions very slowly. Of course, it is possible to formalize this fully. At the same time, it is possible to consider partial behaviour of parameters variations, providing simple scenes of the algorithmic developments.

Let's suppose that the coefficients  $c_j$  of linear criterion function  $Z = \sum_{j=1}^n c_j x_j$  of the linear programming problem

are varying by flow  $B(t, n)$ . Assume that  $t_0$  is the moment where the complete analysis exists, i.e. we know about the existence of optimization at that moment and the optimal vertex and plan, if the latter exists. This vertex obeys the property of stability of optimality for certain variations of coefficients  $c_j$  [GL 1980]. The stability area is described by a set of simple inequalities and it is clear that it is an issue to consider the border of this area. The theoretical analysis of optimality of vertex set elements of the area of base restrictions is well known as the simplex method [V 1998]. The simplex method looks for sequence chains of vertex transitions, which converge to an optimal plan. Complementary, in our case, we study all the possible ways of optimality transitions driven by the changes of coefficients  $c_j$ .

The novelty is that we devise the concept of equivalency of vertices groups of the feasible polyhedron vertices set and prove that the transition from one optimal vertex to another takes place through these groups. So the continuous change of target function coefficients generates the continuous change of optimality vertices.

From practical point of view – a path prediction process is possible to apply to the situation with varying coefficients. Prediction of intersection of the trajectory extrapolation of coefficient changes to the boundary of stability area of the current optimal plan helps to determine the vertex equivalency cluster and so - the further possible transitions and by these – the most likely arriving optimums when coefficients keep the track of their current modifications.

Going through the transitions some vertices might neighbour with comparatively large equivalency groups of vertices and then the total number of those vertices can become large. Theoretically, in terms of flows, having the current optimization vertex, it is necessary to prepare neighbouring equivalent vertices by calculating them by the current and predicted coefficients  $c_j$ . The weakness of the direct application of the given approach is in drastic increase in the number of calculations for the vertex sets involving the predictions and equivalencies. The considered below natural approach gives primary significance to the vertices which correspond to the linear approximations of the given variations.

Let's denote the optimal vertex at the moment  $t_0$  by  $\tilde{x}^{t_0}$  and let  $\tilde{c}^{t_0}$  is the corresponding vector of coefficients.

Let's follow the transition of  $\tilde{c}^{t_0}$  to the  $\tilde{c}^t$ . It is clear that this transition is more or less arbitrary and it is controlled by the flow. It is important if during the transition the vector  $\tilde{c}^t$  of coefficients approaches to the boarder of stability of current optimal plan  $\tilde{x}^{t_0}$ , - or not. To see this we have to monitor the changes of  $\tilde{c}^t$ . Alternatively, it is possible to approximate the transition, activating the possible future “optimal plans”. For example, spline approximations or a more simple approximation by the main values and standard deviations might be applied. The most simple is the linear approximation model, which we consider below. As an extrapolation, it leads to the intersection with the stability boundary (shell) of the vertex  $\tilde{x}^{t_0}$  at the most probability point. In case of sufficient computational resources, it is also possible to consider some neighbourhood of that point, and it is important that in contrast to the above mentioned theoretical model, this applied approach gives an opportunity to work with the limited quantity of the possible candidate vertices. Depending on the considered problem, an algorithmic system is able to choose a corresponding extrapolation scheme, which deals with different numbers of neighbouring vertices. The approximation of  $\tilde{c}^t$  by the flow averages and dispersions requires their calculation, which is a simple flow problem (it is shown in [MM 2003]). Supposing that this question is clarified, let's consider the problem behaviour in the case of linear approximations.

### 3. Linear Approximation

In the case mentioned, variation of the optimization criteria function coefficients is supposed to be changed by an expression  $c_j(\lambda) = c_j^{t_0} + \lambda(c_j^t - c_j^{t_0})$ , where  $\lambda$  varies in certain limits. The interval  $[0,1]$  for  $\lambda$  is internal and characterizes the variation from  $\tilde{c}^{t_0}$  to  $\tilde{c}^t$ , and the values  $\lambda > 1$  are extrapolating the further behaviour of coefficients in a linear model. Let's denote  $c_j^\Delta = c_j^t - c_j^{t_0}$ .

So we are given the linear function

$$(1) \quad Z = \sum_{j=1}^n (c_j^{t_0} + \lambda c_j^\Delta) x_j$$

and the system of linear requirements, given by

$$(2) \quad \begin{aligned} \sum_{i=1}^n a_{ij} x_j &= b_i, \quad i = 1, 2, \dots, m, \\ x_j &\geq 0, \quad j = 1, 2, \dots, n. \end{aligned}$$

It is necessary to accompany the changes of  $\lambda$ , finding out in the interval  $1 < \lambda$  the minimal value at which the change of the optimal plan takes place for the first time. Assume that the vector  $\tilde{x}^t = (x_1^t, x_2^t, \dots, x_n^t)$  which satisfies the system (2) introduces the corresponding new optimization basis.

According to the assumptions, we have optimal solution when  $\lambda = 0$ . Assume that the solution basis consists of the first  $m$  vectors of  $\bar{a}_1, \dots, \bar{a}_n$ . In accord to the simplex algorithm and its optimization condition, all the "estimations" in this case must obey to the following condition:  $z_j - c_j^{t_0} \leq 0, j = 1, 2, \dots, n$ .

As  $c_j(\lambda) = c_j^{t_0} + \lambda c_j^\Delta$  then that general optimization condition becomes:

$$z_j - c_j(\lambda) = (c_j^{t_0} + \lambda c_j^\Delta) x_j^0 - (c_j^{t_0} + \lambda c_j^\Delta) \leq 0, \quad j = 1, 2, \dots, n.$$

Let's group the expression in the following way:

$$c_j^{t_0} (x_j^0 - 1) + \lambda c_j^\Delta (x_j^0 - 1) \leq 0, \quad j = 1, 2, \dots, n,$$

and let introduce the notations:  $\alpha_j = c_j^{t_0} (x_j^0 - 1)$  and  $\beta_j = c_j^\Delta (x_j^0 - 1)$ . The constants  $\alpha_j$  and  $\beta_j$  are defined by the initial configuration: optimization criterion function coefficients and the corresponding solution basis, criterion current coefficients with the supposition that optimization did not change during that period.

For  $\lambda = 0$  we have the optimization vertex  $\tilde{x}^0$ , and therefore, we get the following limitations:

$\alpha_j = c_j^{t_0} (x_j^0 - 1) \leq 0$ . The optimization vertex change does not take place when  $0 \leq \lambda \leq 1$ , so we get also:

$$\alpha_j + \lambda \beta_j = c_j^{t_0} (x_j^0 - 1) + \lambda c_j^\Delta (x_j^0 - 1) \leq 0.$$

In particular, when  $\lambda = 1$  we get  $\alpha_j + \beta_j \leq 0$ . The extreme condition will be written in the following general form:

$\alpha_j + \lambda \beta_j \leq 0, j = 1, 2, \dots, n$ . Let's find the minimal value of  $\lambda$  at which at least one of this inequalities violates for the first time.

Let's separate negative and positive cases of  $\beta_j$ . The restrictions on  $\lambda$  will accept the following forms:

$$\lambda \geq -\alpha_j / \beta_j \text{ for all } \beta_j < 0, \text{ and}$$

$$\lambda \leq -\alpha_j / \beta_j \text{ for all } \beta_j > 0.$$

Let's introduce one more notation:

$$\bar{\lambda} = \left\{ \min(-\alpha_j/\beta_j) \text{ if } \alpha_j > 0, \text{ and } +\infty, \text{ when all } \beta_j \leq 0 \right.$$

The optimal solution for  $\lambda = 0$  coincides with optimal solutions for all  $\lambda$  which obeys the condition  $0 \leq \lambda \leq \bar{\lambda}$ . It is ensued that  $\bar{\lambda}$  is the possible transition configuration. If  $\bar{\lambda} = +\infty$  then there is no change of optimal plan. If  $\bar{\lambda}$  is limited then it is necessary to consider two cases: the first one (a/) is the point  $\bar{\lambda}$  with the possible equivalent optimal plans and possible continuations in this case, and the second one (b/): if there is a new optimal plan and if the problem has no solution at  $\lambda > \bar{\lambda}$ .

a/ Assume that  $\bar{\lambda}$  is finite, i.e.  $\bar{\lambda} = -\alpha_k/\beta_k$  for the corresponding value of parameter  $k$ . It means that  $z_k - c_k(\lambda) = 0$  from which follows that the optimization plan is not single. Actually, let's insert the  $k$ -th vector into the basis and according to the simplex method let's exclude one of the vectors from the previous basis. We will get a new optimal plan the criterion value of which will stay unchanged. It follows even more – that, by all null estimations and by all basis modifications we can get many optimization equivalent vertexes and all elements of their linear closure also have the same discussing optimization value.

b/ In this case, we consider the values  $\lambda > \bar{\lambda}$  and the  $\bar{\lambda}$  is finite. If the coefficients of above mentioned  $k$ -th vector all not positive, i.e.  $\tau_{ik} \leq 0$ , by optimization basis, then according to the simplex method, the criterion function becomes unlimited. This takes place any time when according to the increasing character of the criterion function we get the vector which is going to be involved into the basis  $z_k - c_k(\lambda) > 0$ , but it becomes clear that the vector has no positive  $\tau_{ik} > 0$  coordinate because of we could exclude it from the basis. In this case, it is impossible to choose such a coefficient  $\theta > 0$  that any  $x_i - \theta\tau_{ik} = 0$  when  $i = \{1, \dots, m\}$ . Therefore, we get the optimization plan with  $m+1$  positive components; the set of  $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_m, \bar{a}_k$  vectors are linearly depending and this corresponds to the non-angle vertex. Therefore, linear criterion function could not get to its minimal value. This means that hyper-plane which is defined by linear function could not become supporting hyper-plane of permissible polyhedron at any shift in the direction of gradient.

If a  $\tau_{ik} > 0$  then the vector  $\bar{a}_k$  is included into the basis and another vector  $\bar{a}_l$  is excluded from it. As the new basis is constructed by the simplex method then it corresponds to a new optimal solution, and at those inequalities

$$(3) \quad \alpha'_j + \lambda\beta'_j \leq 0, \quad j = 1, 2, \dots, n$$

are compatible.

Let's show that any  $\lambda < \bar{\lambda}$  does not satisfy the system (3) of inequalities. Really, for the vector  $\bar{a}_l$ , excluded from the basis we will get the following:

$$(4) \quad \alpha'_l = -\alpha_k/\tau_{lk}; \quad \beta'_l = -\beta_k/\tau_{lk},$$

where  $\tau_{lk} > 0$ . Suppose that (3) takes place for any  $\lambda < \bar{\lambda}$  then  $\alpha'_j + \lambda\beta'_j \leq 0$ , or according to (4)  $-\alpha_k - \lambda\beta_k \leq 0$ . As  $\beta_k > 0$  then, from the latter inequality follows that  $\lambda \geq -\alpha_k/\beta_k = \bar{\lambda}$ .

---

#### 4. Conclusion

The paper is devoted to the discussion of applied algorithms for data flows. The linear programming problems and the simplex algorithm of their solution were considered. This research is not about the simplex algorithm developments, but is about the approaches processed in this sphere that also help when according to the problem assumption the coefficients of criterion function variate in the result of the data flows analysis. We got that it is possible to introduce and develop the concepts and tools related to the simplex algorithm by approaches, which solve flow linear optimization problems. The core result is the construction of the extrapolation mechanism that applies linear extrapolation by predicting the stationary data. The concept of equivalency of optimal vertexes

is introduced, which helps to accompany the variation process preparing the possible optimization vertexes in advance.

This is important from the viewpoint of linear programming systems and optimization in applied data flow systems.

---

### Bibliography

- [AJ 2001], M. Ajtai, T. Jayram, R. Kumar, and D. Sivakumar. Counting inversions in a data stream. manuscript, 2001.
- [AC 2003], Aslanyan L., Castellanos J., Mingo F., Sahakyan H., Ryazanov V., Algorithms for Data Flows, International Journal "Information Theories and Applications", ISSN 1310-0513, 2003, Volume 10, Number 3, pp. 279-282.
- [AM 1996], N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In Proc. of the 1996 Annual ACM Symp. on Theory of Computing, pages 20-29, 1996.
- [GL 1980], E.N. Gordeev and V.K. Leontiev, Stability in problems of narrow places, JCM&MP, No. 4, Moscow, 1980.
- [MM 2003], Manoukyan T and Mirzoyan V., Image Sequences and Bit-Plane Differences. "Computer Science & Information Technologies Conference", Yerevan, September 22-26, pp. 284-286 (2003).
- [GG 2002], A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In Proc. of the 2002 Annual ACM Symp. on Theory of Computing, 2002.
- [V 1998], R.J. Vanderbei, Linear Programming; Foundations and Extensions, Kluwer Academic Publishers, Boston/London/Dordrecht, 1998.

---

### Author's Information

**Levon Aslanyan** – Institute for Informatics and Automation Problems, NAS Armenia, P.Sevak St. 1, Yerevan-14, Armenia; e-mail: [lasl@sci.am](mailto:lasl@sci.am)

## THE BOUNDARY DESCRIPTORS OF THE $n$ -DIMENSIONAL UNIT CUBE SUBSET PARTITIONING<sup>1</sup>

**Hasmik Sahakyan, Levon Aslanyan**

**Abstract:** *The specific class of all monotone Boolean functions with characteristic vectors of partitioning of sets of all true-vertices to be minimal is investigated. These characteristic vectors correspond to the column-sum vectors of special  $(0,1)$ -matrices – constructed by the interval bisection method.*

**Keywords:** *monotone Boolean functions,  $(0,1)$ -matrices.*

---

### 1. Introduction

The problem of general quantitative description of vertex subsets of  $n$  dimensional unit cube  $E^n$ , through their partitions, the existence problem and composing algorithms for vertex subsets by the given quantitative characteristics of partitions are considered. Each of these sub-problems has its own theoretical and practical significance. The existence and composing problems are studied intensively [BI, 1988; C, 1986; S, 1986], but the complexity issues are still open [B 1989]. Therefore, studying the problem in different restrictions, for particular type of subsets, and obtaining different necessary and sufficient conditions is essential.

---

<sup>1</sup> The research is supported partly by INTAS: 04-77-7173 project, <http://www.intas.be>

Concerning the problem of general quantitative description - the complete description of the set of all integer-valued vectors - quantitative descriptors of subset partitions is obtained [S 1997]. The description is based on structures, where the characteristic vectors corresponding to the monotone Boolean functions of  $E^n$  have an important role. The main result of this research concerns with the partitioning-boundary-cases, which correspond to the special monotone Boolean functions having minimal characteristic partitioning vectors, which, in their turn, correspond to the column sum vectors of  $(0,1)$ -matrices constructed by the interval bisection method.

## 2. Structure

Let  $M \subseteq E^n$  be a vertex subset of fixed size  $|M| = m$ ,  $0 \leq m \leq 2^n$ . An integer, nonnegative vector  $S = (s_1, s_2, \dots, s_n)$  is called the **characteristic vector of partitions** of set  $M$ , if its coordinates equal to the partition-subsets sizes of  $M$  by coordinates  $x_1, x_2, \dots, x_n$  - the Boolean variables composing  $E^n$ .  $s_i$  equals the size of one of the partition-subsets of  $M$  by the  $i$ -th direction and then  $m - s_i$  is the size of the complementary part of partition. To make this notation precise we will later assume that  $s_i$  is the size of the partition subset with  $x_i = 1$ .

The complete description of all integer-coordinate vectors, which serve as characteristic vectors of partitions for vertex subsets of size  $m$  sets is based on structures, where characteristic vectors corresponding to the monotone Boolean functions play an important role.

Let  $\Xi_{m+1}^n$  denotes the set of all vertices of  $n$  dimensional,  $m+1$  valued discrete cube, i.e. the set of all integer-vectors  $S = (s_1, s_2, \dots, s_n)$  with  $0 \leq s_i \leq m$ ,  $i = 1, \dots, n$ .  $\Psi_m$  denotes the set of all characteristic vectors of partitions of  $m$ -subsets of  $E^n$ . It is evident, that -  $\Psi_m \subseteq \Xi_{m+1}^n$ . Below, the description of  $\Psi_m$  has been given in terms of  $\Xi_{m+1}^n$  geometry: the vertices are distributed schematically on the  $m \cdot n + 1$  layers of  $\Xi_{m+1}^n$  according to their weights - sums of all coordinates. The  $l$ -th layer contains all vectors  $S = (s_1, s_2, \dots, s_n)$  with  $l = \sum_{i=1}^n s_i$ .

Let  $\widehat{\Psi}_m$  and  $\widetilde{\Psi}_m$  are subsets of  $\Psi_m$ , consisting of all its upper and lower boundary vectors, correspondingly:

$\widehat{\Psi}_m$  is the set of all "upper" vectors  $S \in \Psi_m$ , for which  $R \notin \Psi_m$  for all  $R \in \Xi_{m+1}^n$ , greater than  $S$ .  $\widetilde{\Psi}_m$  is the set of all "lower" vectors  $S \in \Psi_m$ , for which  $R \notin \Psi_m$  for all vectors  $R$  from  $\Xi_{m+1}^n$ , less than  $S$ .

These sets of all "upper" and "lower" boundary vectors have symmetric structures - for each upper vector there exists a corresponding (opposite) lower vector, and vice versa; so that also the numbers of these vectors are equal:

$$\widehat{\Psi}_m = \{ \widehat{S}_1, \dots, \widehat{S}_r \} \text{ and } \widetilde{\Psi}_m = \{ \widetilde{S}_1, \dots, \widetilde{S}_r \}.$$

Let  $\widehat{S}_j$  and  $\widetilde{S}_j$  be an arbitrary pair of opposite vectors from  $\widehat{\Psi}_m$  and  $\widetilde{\Psi}_m$  correspondingly.  $I(\widehat{S}_j)$  (equivalently  $I(\widetilde{S}_j)$ ) will denote the minimal sub-cube of  $\Xi_{m+1}^n$ , passing through this pair of vectors. Then,

$$I(\widehat{S}_j) = \{ Q \in \Xi_{m+1}^n / \widehat{S}_j \leq Q \leq \widetilde{S}_j \} \text{ (the coordinate-wise comparison is used).}$$

The following Theorem states that the minimal sub-cubes passing the pairs of corresponding opposite vectors of the boundary subsets are continuously and exactly filling the vector area  $\Psi_m$ .

**Theorem 1** [S 1997]:  $\Psi_m = \bigcup_{j=1}^r I(\widehat{S}_j)$ .

### 3. Boundary Cases

Boundary vectors of  $\psi_m$  can be described by the monotone Boolean functions, defined on  $E^n$ : the set of all characteristic boundary vectors is a subset of the set of all characteristic vectors of partitions of monotone Boolean functions. This fact is confirmed by the following theorem.

**Theorem 2** [S 1997]:  $\widehat{\psi}_m \subseteq M_m^1$  and  $\widetilde{\psi}_m \subseteq M_m^0$ ,

where  $M_m^1$  and  $M_m^0$  are the sets of characteristic vectors of those  $m$ -subsets of  $E^n$ , which correspond to the sets of all one-valued vertices and all zero-valued vertices defined by monotone Boolean functions correspondingly.

Let  $\mathcal{P}_{min}(m, n)$  is the class of monotone Boolean functions for which the partitioning characteristic vectors of  $\widehat{\psi}_m$  are placed on the minimal possible layer of  $\Xi_{m+1}^n$ , - denote this layer by  $L_{min}$ . A similar class of functions is related to  $\widetilde{\psi}_m$ .

The structure of functions of class  $\mathcal{P}_{min}(m, n)$  is related to the linear ordering of vertices of  $E^n$  by decreasing of numeric values of the vectors expressed as  $x_n, x_{n-1}, \dots, x_1$ . Call this sequence of vertices "decreasing", denoting it by  $D_n$ .  $D_n$  has some useful properties. The first  $2^{n-1}$  vertices of  $D_n$  compose an  $n-1$  dimensional sub-cube (interval), incident to the vertex  $x_i = 1, i = \overline{1, n}$ . The reminder  $2^{n-1}$  vertices are in the "shadow" of that interval by the variable  $x_n$ . The vertex order in both sub-cubes correspond to the orderings in  $D_{n-1}$ . In general, for an arbitrary  $k$  the first  $2^k$  vertices of the sequences compose  $k$  dimensional intervals, where variables  $x_n, x_{n-1}, \dots, x_{k+1}$  accept fixed values; and then the next  $2^k$  vertices are in the "shadow" of that interval by the direction  $x_{k+1}$ , and continuing the same way we will receive the recursive structure of the series  $D_n$ .

The given construction provides that the initial segments by  $m$  of the "decreasing" sequence  $D_n$  correspond to the sets of all one-valued-vectors of some monotone Boolean function. We denote this functions by  $\mu(m, n)$ .

Next theorem confirms that the constructed functions are the required monotone Boolean functions.

**Theorem 3:**  $\mu(m, n) \in \mathcal{P}_{min}(m, n)$ .

The general technique to prove this Theorem consists of partitioning of  $E^n$  by one or two directions, the inductions by these parameters and the partitioning of all possible cases into the several sub cases.

The minimal layer  $L_{min}$  might be presented by the following formula:

$$L_{min} = \sum_{i=1}^n s_i = \sum_{i=1}^p \left( (n - k_i - (i - 1)) \cdot 2^{k_i} + k_i \cdot 2^{k_i - 1} \right),$$

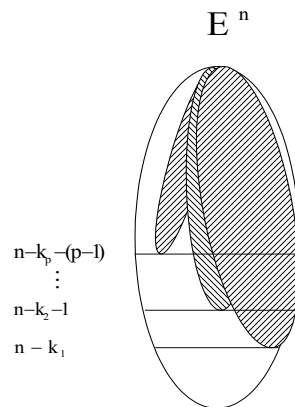
where parameters  $k_i$  correspond to the binary presentation of number  $m$ :

$$m = 2^{k_1} + 2^{k_2} + \dots + 2^{k_p}.$$

Numerical expressions of coordinates of characteristic vector  $S = (s_1, s_2, \dots, s_n)$  of function  $\mu(m, n)$  are also calculable:

$$s_{k_j + 1} = \left( \sum_{l=1}^{j-1} 2^{k_l - 1} \right) + 2^{k_j}, \quad j \in \overline{1, p}$$

for digits  $k_1, k_2, \dots, k_p$  of binary presentation, and



$$s_i = \left( \sum_{l=1}^j 2^{k_l-1} \right) + \left( \sum_{l=j+1}^p 2^{k_l} \right) \text{ for } i, k_{j+1} + 2 \leq i \leq k_j, j \in \overline{1, p-1}$$

and

$$s_i = \sum_{l=1}^p 2^{k_l-1} = m/2 \text{ for } i, 1 \leq i \leq k_p$$

$$s_i = \sum_{l=1}^p 2^{k_l} = m \text{ for } i, k_1 + 2 \leq i \leq n$$

#### 4. Synthesis by Bisections

Subsets  $M \subseteq E^n$  might be presented (coded) by  $(0,1)$ -matrices with  $m$  rows and  $n$  columns, where the rows correspond to the vertices-elements of  $M$ . All the rows are different, and the quantities of 1's in columns (column sums) correspond to the sizes of (one of the two) partition-subsets of  $M$  in corresponding directions.

Thus the existence and synthesis issues of vertex subsets by given quantitative characteristics of their partitions might be reformulated as the corresponding existence and synthesis problems for  $(0,1)$ -matrices.

Let it is given an integer vector  $S = (s_1, \dots, s_n)$ ,  $0 \leq s_i \leq m$ ,  $i = 1, \dots, n$ . If there exists an  $m \times n$   $(0,1)$ -matrix with all different rows and with  $s_i$  column sums, then by the finite number of row pairs replacements it can be transformed into the equivalent matrix of "canonical form" – where each column consists of continuous intervals of 1's and (then - below) 0's such that they compose bisections of intervals of the previous stage (column). Therefore the problem of synthesis of the given  $(0,1)$ -matrices in supposition of the existence might be solved, in particular, by the algorithms which compose the matrices in column-by-column bisection fashion.

The first column is being constructed by allocating of  $s_1$  1's to the first  $s_1$  rows-positions followed by the  $m - s_1$  0's in others. Two intervals is the result: – the  $s_1$  interval of 1's, and the  $m - s_1$  interval of 0's. Within each of these intervals we have the same row, and pairs of  $(i, j)$  rows with  $i$  and  $j$  from different intervals are different.

The second column has been constructed by a similar bi-partitioning of intervals of the first column - situating first 1's and then 0's on these intervals such that the summary length of 1-intervals is equal to  $s_2$ , and the summary length of 0-intervals is equal to  $m - s_2$ . The current  $k$ -th column has been constructed by consecutive and continuous bi-partitioning of intervals of the  $k - 1$  column – providing only that the summary length of all 1's-intervals is equal to  $s_k$ .

When in some column we get all 1 length intervals, then all the rows of matrix become different by the set of constructed columns, and the remainder columns might be constructed arbitrarily.

Partitioning of the intervals in each step can be performed by different ways - following different goals. Let assume that the partitioning of intervals aims to maximize some quantitative characteristics, which might lead to the matrices with different rows. One of such characteristics - the number of pairs of different rows –

is considered in [S, 1995], where is proven that the optimal partitioning is achieved when intervals are partitioned by the equi-difference principle.

In an analogue sophisticated situation when we are not restricted by column sums (or when we allow the whole set of descriptions) and the aim is only in minimization of number of columns for differentiations of rows or in maximization of the number of different row pairs, the best partitioning is known by the "interval bisection" method, which requires the number of columns -  $n = \lceil \log_2 m \rceil$  [K, 1998].

For the given  $m$  let's describe all the possible column sum vectors of matrices composed by the interval bisection method.

In case of  $m = 2^n$  it is evident, that  $s_i = m - s_i = 2^{n-1}$  for  $i = 1, \dots, n$ .

For an arbitrary  $m$ , ( $m > 1$ ), an odd length interval may appear during the separate column partitioning stage. Partitioning the odd length interval takes an extra 1 or 0, which leads to the different column sums. Factually it is satisfactory to consider the case when in each stage an extra 1 is allocated on each odd length interval during the partitioning and let this leads to the column sums  $s_1, s_2, \dots, s_n$ . Column sums corresponding to all the other allocations of extra 1's and 0's can be received through the inversions and monotone transformation of  $s_1, s_2, \dots, s_n$  [S, 1995].

So the number of odd length intervals of each column is equivalent to the difference of 1's and 0's used in the next column, and the goal becomes to calculate the number of odd length intervals (denote it by  $d_i$  for  $i$ -th column).

Let  $2^{n-1} \leq m < 2^n$  and  $m = 2^{k_1} + 2^{k_2} + \dots + 2^{k_p}$  is the binary presentation of number  $m$ , where  $n-1 = k_1 > k_2 > \dots > k_p \geq 0$ .

Since  $m \geq 2^{n-1}$ , then the  $i$ -th column for  $i = 1, \dots, n-1$  will create  $2^i$  intervals. These intervals will have the length approximately equal to

$$2^{k_1-i} + 2^{k_2-i} + \dots + 2^{k_p-i}. \quad (1)$$

The same time, the  $n$ -th column constructs the  $m$  intervals of length 1.

Consider the following cases:

$$k_p - i > 0$$

the  $i$ -th column has no intervals of odd length, therefore  $s_{i+1} = m - s_{i+1} = m/2$ .

$$k_p - i = 0$$

the  $i$ -th column will have  $2^i$  odd length intervals, therefore  $s_{i+1} - (m - s_{i+1}) = 2^i$  and  $s_{i+1} = \frac{m + 2^i}{2}$ .

$$k_1, k_2, \dots, k_j \geq i \text{ and } k_{j+1}, \dots, k_p < i$$

$$\underbrace{2^{k_1-i} + \dots + 2^{k_j-i}}_{\substack{k_j-i \geq 0 \\ (I)}} + \underbrace{2^{k_{j+1}-i} + \dots + 2^{k_p-i}}_{\substack{k_{j+1}-i < 0 \\ (II)}} \quad (2)$$

$$1. \quad k_j - i = 0,$$

Component (I) in (2) provides that each of  $2^i$  intervals of  $i$ -th stage is at least of length  $2^{k_1-i} + \dots + 2^{k_j-i} = q$  (which is odd) and the component (II) - that the length of  $2^{k_{j+1}-i} + \dots + 2^{k_p-i}$  share of  $2^i$  intervals is  $q+1$ . Therefore the number of even intervals equals

$\left( \sum_{l=j+1}^p 2^{k_l-i} \right) \cdot 2^i = \sum_{l=j+1}^p 2^{k_l}$  and the number of odd length intervals is equal to

$$d_i = 2^i - \sum_{l=j+1}^p 2^{k_l} = 2^{k_j} - \sum_{l=j+1}^p 2^{k_l}$$



$$\text{Hence: } s_{i+1} = \frac{\sum_{l=1}^p 2^{k_l} + 2^{k_j} - \sum_{l=j+1}^p 2^{k_l}}{2} = \left( \sum_{l=1}^{j-1} 2^{k_l-1} \right) + 2^{k_j} \quad (3)$$

for  $i = k_j$  and the same formula holds for each  $k_j, j = 1, 2, \dots, p$ .

2.  $k_j - i > 0$ ,

in this case component (I) in (2) provides that each of  $2^i$  intervals of  $i$ -th stage is of even length and the component (II) - that the  $2^{k_{j+1}-i} + \dots + 2^{k_p-i}$  share of  $2^i$  intervals is of odd length.

The number of odd length intervals is equal to

$$d_i = \left( \sum_{l=j+1}^p 2^{k_l-i} \right) \cdot 2^i = \sum_{l=j+1}^p 2^{k_l} \quad \text{and} \quad s_{i+1} = \frac{\sum_{l=1}^p 2^{k_l} + \sum_{l=j+1}^p 2^{k_l}}{2} = \left( \sum_{l=1}^j 2^{k_l-1} \right) + \left( \sum_{l=j+1}^p 2^{k_l} \right) \quad (4)$$

for  $i \neq k_j$  and the same formula holds for each  $k_j, j = 1, 2, \dots, p$ .

Thus (3) and (4) describe column sums of matrices composed by interval bisection method by the given  $m$ . Comparing this column sums to the coordinates of characteristic vector of monotone Boolean function  $\mu(m, n)$  points out that the (0,1)-matrix  $A$ , which corresponds to the set of all true-vectors of monotone Boolean function  $\mu(m, n)$  has the same set of column sums as the matrix – constructed by the bisection method.

---

## Conclusion

---

Resuming, - the n-cube subsets and partitioning (Set Systems and characterization by inclusion of a particular element) in specific boundary cases, and the bisection strategy characterization are strongly similar having the same characteristic numerical descriptors, simply related to the binary representations of the set sizes.

---

## Bibliography

---

- [B, 1989] C. Berge, Hypergraphs. North Holland, Amsterdam, 1989.
- [BI, 1988] Billington D., Conditions for degree sequences to be realisable by 3-uniform hypergraphs". The Journal of Combinatorial Mathematics and Combinatorial Computing". 3, 1988, 71-91.
- [C, 1986] Colbourn Charles J., Kocay W.L. and Stinson D.R., Some NP-complete problems for hypergraph degree sequences. Discrete Applied Mathematics 14, p. 239-254 (1986).
- [K, 1998] D. Knuth, The Art of Computer Programming, second edition, vol.3. Sorting and Searching, Addison-Wesley, 1998
- [R, 1966] H. J. Ryser. Combinatorial Mathematics, 1966.
- [S, 1986] H. A. Sahakyan. Greedy Algorithms for Synthesis of (0,1) Matrices, Doklady Akademii Nauk Arm. SSR, v. 83, 5, pp. 207-209 (1986)
- [S, 1997] H. Sahakyan. On a class of (0,1)-matrices connected to the subsets partitioning of  $E^n$ , Doklady NAS Armenia, v. 97, 2, pp. 12-16.
- [S, 1995] H. Sahakyan. Hierarchical Procedures with the Additional Constraints, II Russia, with participation of NIS Conference, Pattern Recognition and Image Analysis: New Information Technologies, Ulianovsk, 1995, pp.76-78.
- [SA, 2001] L. Aslanyan, H. Sahakyan. On the boundary cases of partitioning of subsets of the n-dimensional unit cube, Computer Science & Information Technologies Conference, Yerevan, September 17-20, 2001, 164-166.

---

## Authors' Information

---

**Hasmik Sahakyan** – Institute for Informatics and Automation Problems, NAS Armenia, P.Sevak St. 1, Yerevan-14, Armenia; e-mail: [hasmik@ipia.sci.am](mailto:hasmik@ipia.sci.am)

**Levon Aslanyan** – Institute for Informatics and Automation Problems, NAS Armenia, P.Sevak St. 1, Yerevan-14, Armenia; e-mail: [lasl@sci.am](mailto:lasl@sci.am)

## UPDATE-RETRIEVE PROBLEM: DATA STRUCTURES AND COMPLEXITY

Jose Joaquin Erviti, Adriana Toni

**Abstract:** Let  $V$  be an array storing values in an arbitrary commutative semi-group. The update-retrieve problem concerns the study and design of data structures for implementing the following operations. The operation  $update(j,x)$  has the effect  $[v_i \leftarrow (v_i + x)]$ , and the operation  $retrieve(j)$  returns the sum  $\sum_{i \in T_j} v_i$ , being  $T_j$

( $1 \leq j \leq m$ ) non empty subsets of  $\{1, 2, \dots, n\}$ . These tasks are to be performed on-line. Different data structures involving different number of variables may be used to solve this computational problem, and the complexity of the operations will depend on the choice. In this paper we work inside an algebraic model of computation. Data structures are defined within the model, and the complexity will be measured relative to it.

**Keywords:** data structures, models of computation, analysis of algorithms and problem complexity

### Introduction

Let  $S$  be an arbitrary commutative semi-group and let  $V=(v_1 \dots v_n)$  be an array of variables which assume values in  $S$ . Let  $T_1, T_2, \dots, T_m$  be non-empty subsets of  $\{1, 2, \dots, n\}$ . Our interest centers upon the study and design of data structures for representing  $V$  assuming that we desire to execute the following operations on the variables  $v_1 \dots v_n$ :

- a) Retrieve( $j$ ): return  $\sum_{i \in T_j} v_i \quad \forall 1 \leq j \leq m$
- b) Update( $i,x$ ):  $v_i := v_i + x \quad \forall 1 \leq i \leq n, x \in S$  (0)

A specific instance of (0) obtained by specifying  $n, m$  and  $T_1, T_2, \dots, T_m$ , represents a computational problem, which from now on will be referred to as *update retrieve problem of size  $(m,n)$* , abbr. UR- $(m,n)$ . If these quantities are left unspecified, then (0) defines an ensemble of such problems.

If for example the sets  $T_j$  are defined as

$T_j = \{1, 2, \dots, j\}$ ,  $1 \leq j \leq n$ , then the problem is known as the *array maintenance problem of order  $n$* , abbr. AMP- $n$ . This problem is well understood in the semi-group model of computation (that is, assuming that the array stores values in a commutative semi-group).

A different computational problem is given by choosing the collection of sets  $T_j$  to comprise the intervals  $\{r, r+1, \dots, s\}$ ,  $1 \leq r \leq s \leq n$ , and in this case we are dealing with the *the range query problem*.

A matricial model of computation has been defined for the study of the UR- $(m,n)$  problem, no matter the choice of the sets  $T_1, T_2, \dots, T_m$  (see [6]). The model generalizes the algebraic model of computation defined in [3] for the AMP- $n$ , and the one defined in [5] for the study of the range query problem, relative to which computational complexity is assessed, in order to accommodate the whole class of problems described in (0).

The model defined in [6] consists of triples  $\langle R, U, Z \rangle$  where

$Z = \{z_1 \dots z_n\}$  is a set of variables,  $R=(r_{i,j})$  is an  $m \times n$  matrix and  $U=(u_{i,j})$  is an  $m \times n$  matrix -  $m, n$  are the parameters specifying an UR- $(m,n)$  problem. Each row of  $R$  describes the subset of variables of  $Z$  which have to be added to execute one of the Retrieve operations, and the  $i$ -th column of  $U$  describes the subset of such variables which have to be incremented to execute Update( $i$ ). So, a pair of  $R$  and  $U$  matrices describes a solution for an update retrieve problem of size  $(m,n)$ .

Associated with a triple  $\langle R, U, Z \rangle$ , the programs implementing the operations are defined as:

- 1) Update(j,x): for k=1 until l do  $z_k := z_k + u_{k,j} x$
- 2) Retrieve(j): output  $\sum_{k=1}^l r_{f(j),k} z_k$  (1)

(f(j) denotes the row of R corresponding to the Retrieve(j) operation, that is to say, the row is associated with the  $T_j$  set)

To accommodate arrays which store values in arbitrary commutative semi-groups, the matrices R and U are defined over the integers.

Let  $R_{i,*}$  denote the  $i$ -th row of matrix R.

In [6] has been proved that the programs in (1) are correct if and only if  $R_{f(j),*} \times U = (w_1 \dots w_n)$ ,  $j=1 \dots m$ , where f is a one to one mapping

$$f: \{1 \dots m\} \rightarrow \{1 \dots m\} \quad \text{and} \quad w_i = \begin{cases} 1 & i \in T_j \\ 0 & \text{otherwise} \end{cases}$$

Within the model, the complexity of an Update(j,x) operation is defined as the number of non-zero entries in the  $j$ -th column of U, and the complexity of a Retrieve(j) operation is defined as the number of non-zero entries in the  $f(j)$ -th row of R.

The next figure shows the product matrix for the AMP-4 ( $T^4$ ) and the range query problem of size 4 ( $H^4$ ) respectively:

$$T^4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad H^4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A *balanced update retrieve problem of size n*, abbr. BUR-n, is any instance of the ensemble of problems described in (0), verifying that  $m=n$ . In [2] has been proved that given any BUR-n, there exists for each  $k \leq n$  a solution such that Updates (respectively Retrieves) take time  $\leq k$  and Retrieves (respectively Updates) take time  $\leq \lceil n / (\log_2 k + 1) \rceil$  (see Theorem 4 in [2]).

In this paper we extend this result presenting solutions for the range query problem within the matricial model defined in [6] which give us worst case complexity less or equal k for the Updates and  $\leq \lceil n / (2\sqrt{k} - 1) \rceil$  for the Retrieves. Let us observe that in this case, the number of sets  $T_1, T_2, \dots, T_m$  is  $n(n+1)/2$ . We calculate as well the number of program variables  $z_1 \dots z_l$  required for these particular solutions.

---

### Solutions for the Range Query Problem of Size n

---

We choose the collection of sets  $T_j$  to comprise the intervals  $\{r, r+1, \dots, s\}$ ,  $1 \leq r \leq s \leq n$ .

Recall that given a triple  $\langle R=(r_{i,j})_{i=1 \dots m, j=1 \dots l}, U=(u_{i,j})_{i=1 \dots l, j=1 \dots n}, Z=\{z_1 \dots z_l\} \rangle$  within the model, the Update and Retrieve operations are implemented as:

1) Update(j,x): for k=1 until l do  $z_k := z_k + u_{k,j} x$

2) Retrieve(j): output  $\sum_{k=1}^l r_{f(j),k} z_k$

(f(j) denotes the row of R corresponding to the Retrieve(j) operation, that is to say, the row is associated with the  $T_j$  set, see (1) within the Introduction section ).

We refer to n as the size of the range query problem.

Our interest centers upon improving the complexity of the operations. Observe that there are  $n + \binom{n+1}{2}$

operations, namely the n operations Update(j,\_) with  $j=1 \dots n$ , and the  $\binom{n+1}{2}$  operations Retrieve(j) with

$T_j = \{r, r+1, \dots, s\}$  for certain  $1 \leq r \leq s \leq n$ .

We work inside the *semi-group model of computation*: the program variables store values from an arbitrary commutative semi-group, and the implementations of the Update and Retrieve operations must perform correctly irrespective of the particular choice of the semi-group. In particular, the implementations are not permitted to utilize the subtraction operation.

Next we define a solution for the range query problem of size n such that the worst case complexity is k for the Updates and  $\leq n/(2\sqrt{k}-1)$  for the Retrieves.

It is based on the solution for a BUR-n described in [2], but now we have  $T_1, T_2, \dots, T_m$  sets with  $m = \binom{n+1}{2}$ ,

and we use less program variables, obtaining a slightly better upper bound for the Retrieves.

Let  $d=2\sqrt{k}-1$  and  $S_j = \{(j-1)d+1, \dots, jd\} \quad \forall j=1 \dots n/d$ .

We define the triple  $\langle R=(r_{i,j}), U=(u_{i,j}), Z \rangle$  as:

- 1)  $Z = \{z_\alpha / \exists i,j. (\alpha \subseteq S_i \wedge \alpha = T_j)\}$
- 2)  $r_{f(j)z} \alpha = \begin{cases} 1 & \exists i, j. \alpha = T_j \cap S_i \\ 0 & \text{otherwise} \end{cases}$
- 3)  $u_{z} \alpha_j = \begin{cases} 1 & j \in \alpha \\ 0 & \text{otherwise} \end{cases} \quad (2)$

Next we show the solution for the range query problem of size n=6 when we choose k=4.

We have  $d=2\sqrt{k}-1$  and so  $d=3$ ,  $S_1 = \{1,2,3\}$ ,  $S_2 = \{4,5,6\}$ .

The set of variables is:

$$Z = \{z^\alpha / \alpha \in \{\{1\}, \{1,2\}, \{1,2,3\}, \{2\}, \{2,3\}, \{3\}, \{4\}, \{4,5\}, \{4,5,6\}, \{5\}, \{5,6\}, \{6\}\}\}$$

And the operations would be implemented as (we explicitly enumerate the elements of each  $T_j$  as the subindex for the corresponding Retrieve operation):

Retrieve<sub>{1}</sub> : output  $z_{\{1\}}$   
 Retrieve<sub>{1,2}</sub> : output  $z_{\{1,2\}}$   
 Retrieve<sub>{1,2,3}</sub> : output  $z_{\{1,2,3\}}$   
 Retrieve<sub>{1,2,3,4}</sub> : output  $z_{\{1,2,3\}}+z_{\{4\}}$   
 Retrieve<sub>{1,2,3,4,5}</sub> : output  $z_{\{1,2,3\}}+z_{\{4,5\}}$   
 Retrieve<sub>{1,2,3,4,5,6}</sub> : output  $z_{\{1,2,3\}}+z_{\{4,5,6\}}$   
 Retrieve<sub>{2}</sub> : output  $z_{\{2\}}$   
 Retrieve<sub>{2,3}</sub> : output  $z_{\{2,3\}}$   
 Retrieve<sub>{2,3,4}</sub> : output  $z_{\{2,3\}}+z_{\{4\}}$   
 Retrieve<sub>{2,3,4,5}</sub> : output  $z_{\{2,3\}}+z_{\{4,5\}}$   
 Retrieve<sub>{2,3,4,5,6}</sub> : output  $z_{\{2,3\}}+z_{\{4,5,6\}}$   
 Retrieve<sub>{3}</sub> : output  $z_{\{3\}}$   
 Retrieve<sub>{3,4}</sub> : output  $z_{\{3\}}+z_{\{4\}}$   
 Retrieve<sub>{3,4,5}</sub> : output  $z_{\{3\}}+z_{\{4,5\}}$   
 Retrieve<sub>{3,4,5,6}</sub> : output  $z_{\{3\}}+z_{\{4,5,6\}}$   
 Retrieve<sub>{4}</sub> : output  $z_{\{4\}}$   
 Retrieve<sub>{4,5}</sub> : output  $z_{\{4,5\}}$   
 Retrieve<sub>{4,5,6}</sub> : output  $z_{\{4,5,6\}}$   
 Retrieve<sub>{5}</sub> : output  $z_{\{5\}}$   
 Retrieve<sub>{5,6}</sub> : output  $z_{\{5,6\}}$   
 Retrieve<sub>{6}</sub> : output  $z_{\{6\}}$

Update(1,x) :  $z_{\{1\}} := z_{\{1\}}+x, z_{\{1,2\}} := z_{\{1,2\}}+x, z_{\{1,2,3\}} := z_{\{1,2,3\}}+x$   
 Update(2,x) :  $z_{\{1,2\}} := z_{\{1,2\}}+x, z_{\{1,2,3\}} := z_{\{1,2,3\}}+x, z_{\{2\}} := z_{\{2\}}+x, z_{\{2,3\}} := z_{\{2,3\}}+x$   
 Update(3,x) :  $z_{\{1,2,3\}} := z_{\{1,2,3\}}+x, z_{\{2,3\}} := z_{\{2,3\}}+x, z_{\{3\}} := z_{\{3\}}+x$   
 Update(4,x) :  $z_{\{4\}} := z_{\{4\}}+x, z_{\{4,5\}} := z_{\{4,5\}}+x, z_{\{4,5,6\}} := z_{\{4,5,6\}}+x$   
 Update(5,x) :  $z_{\{4,5\}} := z_{\{4,5\}}+x, z_{\{4,5,6\}} := z_{\{4,5,6\}}+x, z_{\{5\}} := z_{\{5\}}+x, z_{\{5,6\}} := z_{\{5,6\}}+x$   
 Update(6,x) :  $z_{\{4,5,6\}} := z_{\{4,5,6\}}+x, z_{\{5,6\}} := z_{\{5,6\}}+x, z_{\{6\}} := z_{\{6\}}+x$

### **Proposition 1**

Let the triple  $\langle R=(r_{i,j}), U=(u_{i,j}), Z \rangle$  be defined as in (2). Then  $R \times U = H^n$ , being  $H^n$  the matrix product for the range query problem of size  $n$ .

### **Proof**

The proof is trivial from definition of matrix  $H^n$  and definitions in (2).

### **Proposition 2**

Let the triple  $\langle R=(r_{i,j}), U=(u_{i,j}), Z \rangle$  be defined as in (2). Then the number of program variables required by this

solution is  $|Z| = n \left( \frac{\log_2 k + 2}{2} \right)$

### **Proof**

It is quite easy to verify that for each set  $S_j$  the number of variables is  $\lceil d(d+1)/2 \rceil$ , and the result follows from the fact that  $d$  is defined as  $d=2\sqrt{k}-1$  and there are  $n/d$  sets.

### **Lemma 3**

Let the triple  $\langle R=(r_{i,j}), U=(u_{i,j}), Z \rangle$  be defined as in (2). Then the worst case complexity for any Update operation is  $k$  and the worst case complexity for any Retrieve operation is  $\leq \lceil n/(2\sqrt{k}-1) \rceil$ .

### **Proof**

The result is trivially obtained from the definitions in (2).

---

**Bibliography**

---

- [1] D.J. Volper, M.L. Fredman, *Query Time Versus Redundancy Trade-offs for Range Queries*, Journal of Computer and System Sciences 23, (1981) pp.355--365.
- [2] W.A. Burkhard, M.L. Fredman, D.J.Kleitman, *Inherent complexity trade-offs for range query problems*, Theoretical Computer science, North Holland Publishing Company 16, (1981) pp.279--290.
- [3] M.L. Fredman, *The Complexity of Maintaining an Array and Computing its Partial Sums*, J.ACM, Vol.29, No.1 (1982) pp.250--260.
- [4] A. Toni, *Lower Bounds on Zero-one Matrices*, Linear Algebra and its Applications, 376 (2004) 275--282.
- [5] A. Toni, *Matricial Model for the Range Query Problem and Lower Bounds on Complexity*, submitted.
- [6] A. Toni, *Matricial Model for the Update-retrieve Problem and Upper Bounds on Complexity*, submitted.

---

**Authors' Information**

---

Jose Joaquin Erviti – [jerviti@fi.upm.es](mailto:jerviti@fi.upm.es)

Adriana Toni – [atoni@fi.upm.es](mailto:atoni@fi.upm.es)

Facultad de Informática, Universidad Politécnica de Madrid, Spain

## THE DISTRIBUTED SYSTEM OF DATABASES ON PROPERTIES OF INORGANIC SUBSTANCES AND MATERIALS

**Nadezhda N.Kiselyova, Victor A.Dudarev, Ilya V.Prokoshev, Valentin V.Khorbenko,  
Andrey V.Stolyarenko, Dmitriy P.Murat, Victor S.Zemskov**

**Abstract:** *The principles of organization of the distributed system of databases on properties of inorganic substances and materials based on the use of a special reference database are considered. The last includes not only information on a site of the data about the certain substance in other databases but also brief information on the most widespread properties of inorganic substances. The proposed principles were successfully realized at creation of the distributed system of databases on properties of inorganic compounds developed by A.A.Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences.*

**Keywords:** *database, distributed information system, inorganic substances and materials, reference database.*

---

**Introduction**

---

Now hundreds thousand of inorganic compounds are known. Every year thousands of new substances are added to them. In connection with diversity of applications of inorganic materials, the information on them is scattered over the most various publications. Therefore a search for the information about properties of inorganic compounds, especially if they have been synthesized recently, frequently makes a considerable difficulty and not always it achieves success. A consequence of it is the duplication of investigations on synthesis and research of inorganic substances. In addition, the experts not always can find already synthesized substance that is the most suitable for certain applications. The necessity of acceleration of researches on development and application of new materials were the reasons of creation of numerous databases (DB) on properties of inorganic substances. Thousand such databases considerably have improved information service for the experts in the field of inorganic chemistry and materials science, however there was another problem - problem of search for DB, in which the needed information on the certain inorganic substances is stored.

## Structure of the Distributed System of Databases on Properties of Inorganic Substances and Materials

One of ways of the solution of this problem is the development of some reference database (RDB), which would store the information on where to search for the necessary information on substances. The distributed system of databases of A.A.Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences (IMET RAS) (fig.1), submitted in the present paper, is a prototype of such information system. In this case the role of a reference database carries out the DB on properties of inorganic substances «Phases» [Kiseleva et al., 1996] which contains not only data on a site of the various information in other DB but also brief information on the most widespread properties of tens thousand of compounds, for example, melting and boiling points, symmetry of a crystal lattice, etc. (fig.2). RDB provides search for the relevant information on chemical substances and their properties. The detailed information on substances, which have practical importance, is stored in DBs, for example, in DBs on properties of materials for electronics [Kiselyova et al., 2004; Khristoforov et al., 2001] developed by us. Thus, the distributed information system integrated at a level of Web-interfaces is created.

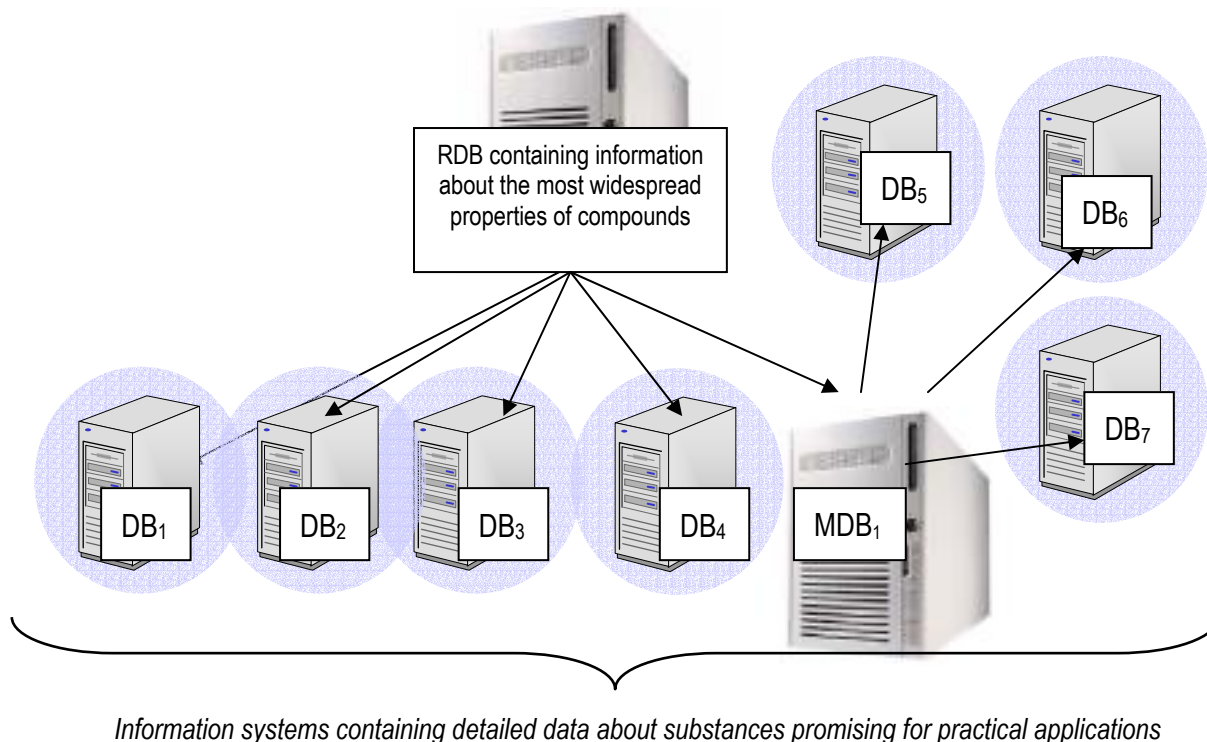


Fig.1. Principles of design of the distributed system of databases on properties of inorganic substances and materials

The concept of design of a special reference database - metabase (MDB) - in the distributed system of the Russian databases on materials for electronics is considered in the terms of the set theory in the paper [Kornyushko et al., 2005]. As shown in this work the search for the relevant information on certain system  $s$  can be reduced to definition of the relation  $R$  being a subset of Cartesian product,  $S \times S$  (in other words,  $R \subset S^2$ ). Here set  $S$  is information on substances and systems stored in MDB. The relation  $R$  is symmetric at design of the distributed system of DBs on materials for electronics [Kornyushko et al., 2005], since the information of integrated DB mutually supplements each other. Let's note that not always relation  $R$  should be strictly symmetric. For example, RDB on properties of inorganic substances "Phase" [Kiseleva et al., 1996] contains only brief information on tens thousand of chemical compounds, but specialized DB on properties

of acoustic-optical, electro-optical and non-linear optical substances “Crystal” [Kiselyova et al., 2004] and DB on the phase diagrams of semiconductor systems “Diagram” [Khristoforov et al., 2001] contain the detailed information on hundreds substances promising for practical applications. Certainly, the users of the specialized systems own more detailed data in comparison with the information stored in RDB “Phase”. Hence, the users of RDB “Phase”, who search for the relevant information, must have an access to the data in specialized DB, and users, for example, of DB “Diagram” do not have the access to the relevant information on properties of compound from RDB “Phase”.

Hence, in this case relation of relevance, given in [Kornyushko et al., 2005], requires the certain updating. Assume that we have relation  $N$ , which describes inadmissible transitions from determined DB into others DB (transitions of a kind  $d_1 \rightarrow d_2$ ). Here set  $D$  is information on databases. That is if  $d_1, d_2 \in D$  and pair  $(d_1, d_2) \in N$ , the transition from integrated information system  $d_1$  into system  $d_2$  is inadmissible.

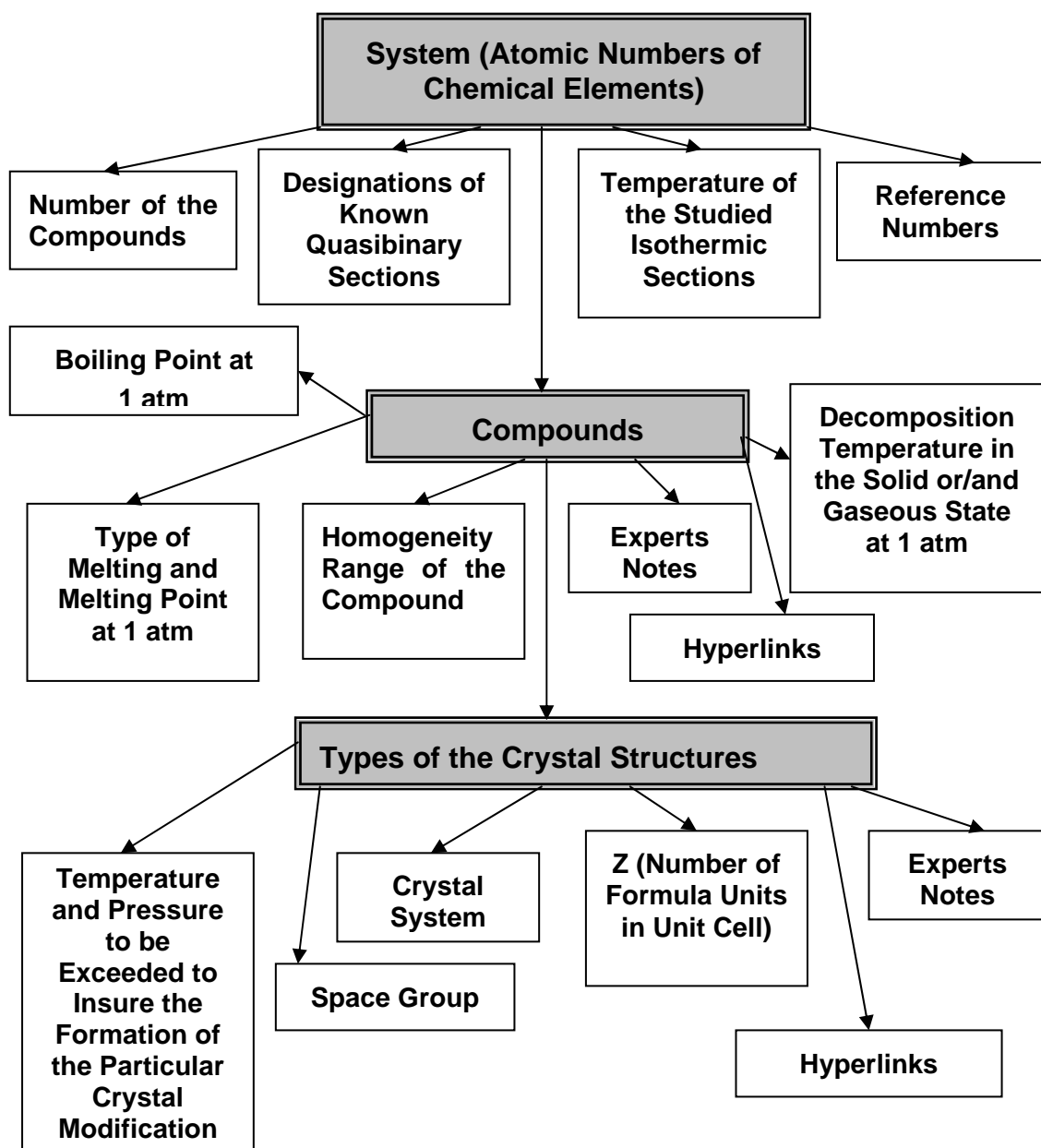


Fig.2. Structure of RDB «Phases»



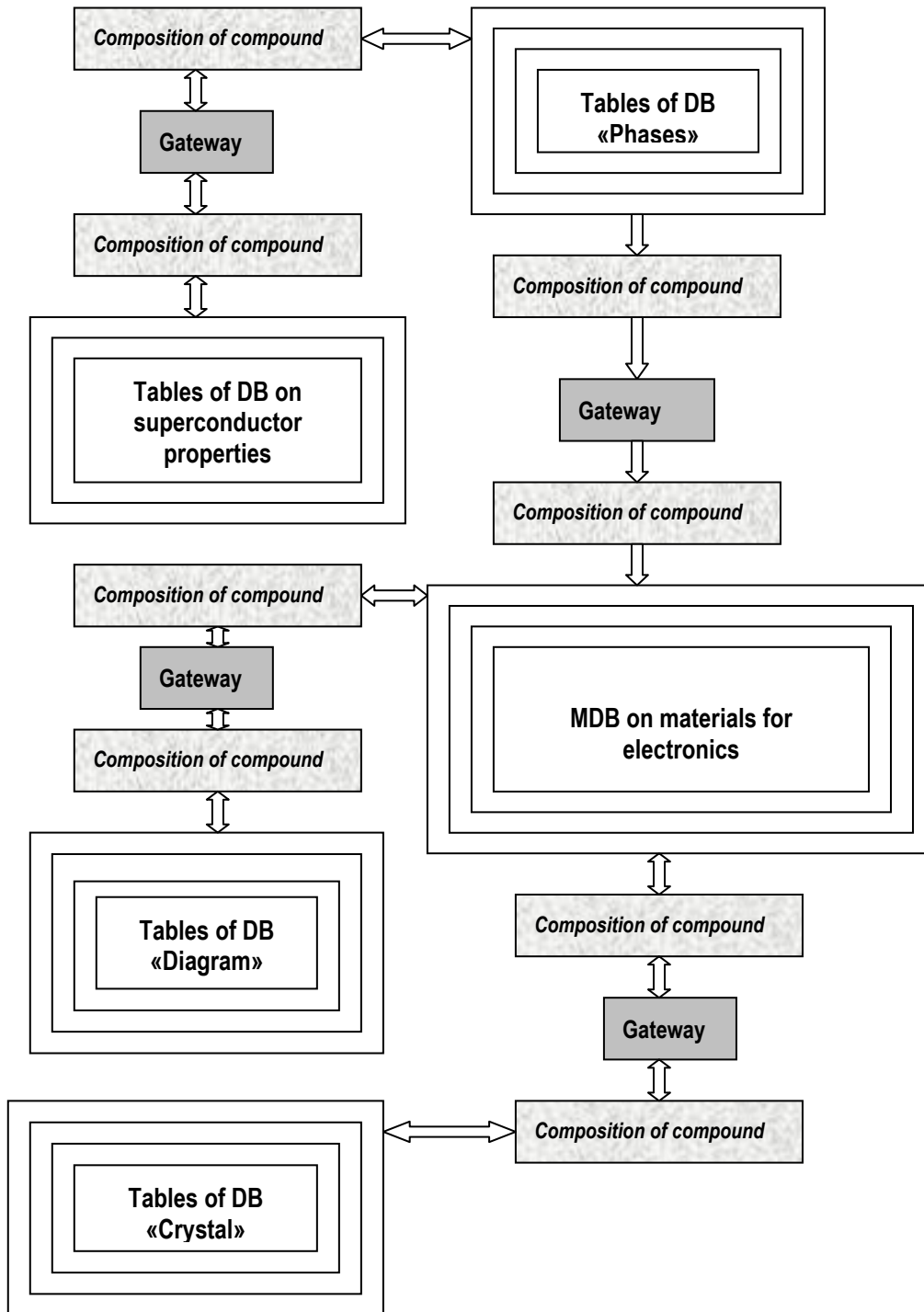


Fig.3. Structure of the distributed system of DBs of IMET RAS

For example, if user looks through the information on certain property of compound in one of DB (i.e. actually there is an access to the information determined by a pair  $(d_1, s_1)$ ), he can have the relevant information on some property of chemical system from another DB, determined by pair  $(d_2, s_2)$ . As a result, user receives the required new relation of relevance  $RN$  as  $RN \subset (d_1, s_1) \times (d_2, s_2)$ , where  $d_1, d_2 \in D; s_1, s_2 \in S$ .

Thus, the new relation of relevance  $RN$  can be constructed on the basis of the old relation  $R$  and set  $N$  according to the following rule: for any chemical systems  $s_1 \in S, s_2 \in S$ , if  $(s_1, s_2) \in R$  and  $d_1, d_2 \notin D$ , then  $(d_1, s_1), (d_2, s_2) \in R$ .

Such decision of a problem of the search for relevant data about properties of substances has many advantages main of which are: simplicity of expansion of the distributed information system, independence on software and hardware platforms, opportunity of actualization and administration of DBs by different organizations which are located in different cities and even in different countries, reduction of traffic, an use of not so powerful, inexpensive servers.

The data on chemical composition (the list of chemical elements and their ratios) are external keys of RDB and various DBs in the distributed system of databases of IMET RAS on properties of inorganic substances and materials (fig.3). It is the most general characteristic of substances, which is inherent in all inorganic objects. Now the distributed system includes besides DB «Phase», in which now the information on properties of ternary compounds is stored, the DB on properties of ternary compounds-superconductors, the DB on properties of acoustic-optical, electro-optical and non-linear optical substances "Crystal" [Kiselyova et al., 2004] and DB on phase diagrams of semiconductor systems "Diagram" [Khristoforov et al., 2001]. The latest two DBs, functioning with the use of various software and hardware platforms, were integrated on the basis of the use of special metabase on properties of materials for electronics [Kiselyova et al., 2004; Korniyushko et al., 2005]. Further the distributed system of databases will include the new DB on bandgaps of semiconductors "BandGap" and also other Russian DBs on properties of materials for electronics: DB on intermolecular potentials for components of the CVD processes in microelectronics (Joint Institute of High Temperature of the Russian Academy of Sciences), information system for modeling processes of preparation of epitaxy of hetero-structures of semiconductor materials by the method of liquid epitaxy (M.V.Lomonosov Moscow State Academy of Fine Chemical Technology), etc.

---

## Conclusion

---

The system of databases of IMET RAS is accessible for the registered users of the Internet:  
<http://www.imet-db.ru>.

The work is supported by RFBR, grant № 04-07-90086.

---

## Bibliography

---

- [Khristoforov et al., 2001] Yu.I.Khristoforov, V.V.Khorbenko, N.N.Kiselyova et al. Internet-accessible database on phase diagrams of semiconductor systems. Izvestiya VUZov. Materialy elektron.tekhniki, 2001, №4 (Russ.).
- [Kiseleva et al., 1996] N.N.Kiseleva, N.V.Kravchenko, and V.V.Petukhov. Database system on the properties of ternary inorganic compounds (IBM PC version). Inorganic Materials, 1996, v.32.
- [Kiselyova et al., 2004] N.N.Kiselyova, I.V.Prokoshev, V.A.Dudarev, et al. Internet-accessible electronic materials database system. Inorganic materials, 2004, v.42, №3.
- [Korniyushko et al., 2005] V.Korniyushko, V.Dudarev. Software development for distributed system of Russian databases on electronics materials. Int. J. "Information Theories and Applications", 2005, v.10.

---

## Authors' Information

---

**Nadezhda N.Kiselyova** – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, senior researcher, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: [kis@ultra.imet.ac.ru](mailto:kis@ultra.imet.ac.ru)

**Victor A.Dudarev** – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, programmer, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: [vic@osg.ru](mailto:vic@osg.ru)

**Ilya V.Prokoshev** – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, programmer, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: [eldream@e-music.ru](mailto:eldream@e-music.ru)

**Valentin V.Khorbenko** – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, programmer, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: [Khorbenko\\_v@mail.ru](mailto:Khorbenko_v@mail.ru)

**Andrey V.Stolyarenko** – Moscow Institute of Electronics and Mathematics (Technical University), post-graduate student, P.O.Box: 109028, B.Trehsvjatitelsky per. 3/12, Moscow, Russia, e-mail: [stol-drew@yandex.ru](mailto:stol-drew@yandex.ru)

**Dmitriy P.Murat** – Moscow Institute of Electronics and Mathematics (Technical University), post-graduate student, P.O.Box: 109028, B.Trehsvjatitelsky per. 3/12, Moscow, Russia, e-mail: [mr\\_wire@mail.ru](mailto:mr_wire@mail.ru)

**Victor S.Zemskov** – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, head of Laboratory of Semiconducting Materials, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: [zemskov@ultra.imet.ac.ru](mailto:zemskov@ultra.imet.ac.ru)

## SOFTWARE DEVELOPMENT FOR DISTRIBUTED SYSTEM OF RUSSIAN DATABASES ON ELECTRONICS MATERIALS

**Valery Kornyshko, Victor Dudarev**

**Abstract:** *Current state of Russian databases on substances and materials properties was considered. It was prepared a brief review of integration methods of given information systems and a distributed databases integration approach was proposed that based on metabase. Implementation details were mentioned on the posed database on electronics materials integration approach. It was considered an operating pilot version of given integrated information system implemented at IMET RAS.*

**Keywords:** *distributed database integration, metabase, Web services, database on electronics materials.*

---

### Introduction

---

Development and utilization of databases on substances and materials properties is a basis in providing information service for specialists in chemistry and materials science. Every research organization aimed at its own data center creation. Such data centers contained information closely related to research areas of particular organization. Historically several data centers were formed for data storage and processing in every organization (scientific research institute or university). This can be explained not only by administrative reasons, but rather by significant differences in problem domain. Existing situation creates great problems for accessing such data, because this information is dispersed over numerous data sources.

At present time, period of such an information fragmentation is coming to the end due to rapid IT-industry development. Present-day progress in science and technique stimulates concentration of diverse information on physicochemical substance properties. Modern polyfunctional materials development requires from us high standard of knowledge in different properties of substances. Efficient online information service (for materials science engineers and chemists providing full data from reliable sources) decreases baseless papers' duplication and ultimately it reduces cost and time required creating modern materials. Inaccessibility and frequently dispersion of information over different heterogeneous data sources makes great difficulties in decision-making process considering application of one or another material.

During development of presented in this paper integrated information system key task was creation of intelligent, simple in architecture and effective software infrastructure. This software infrastructure should integrate data on substances and materials rationally and reasonably. Integration means are required that should be capable to provide not only unified access to operating data centers, but these integration means should allow us to create comprehensive data access infrastructure that is based on unified standards and uniform network interconnection principles also.

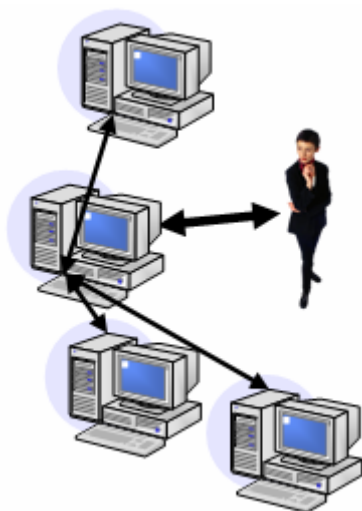
## Database Integration Approaches Overview

Principally there are two approaches to database integration.

The first one implies full merging of existing resources. That is the case when database complex is a single information system (megabase) for end users, operators and administrators. Database exploitation costs reduction and information duplication decrease can be mentioned among advantages of such variant.

Every data center is a point of information concentration and online data analytical processing. In addition, technology of information accumulating and data processing is settled down in each organization. Moreover, great investments that were made in hardware and software do not allow solving the data dissociation problem by all data mechanical transportation to some centralized database. Moreover Russian databases on electronics materials were developed in various organizations and thus they took advantage of different database management systems (DBMS). Taking into consideration differences in data quality, data expertise, data store types and many other troubles emerging when changing existing systems operating principles it should be stated that full and smooth integration is practically impossible for above mentioned resources.

The second integration approach main essence is that we are not going to integrate databases themselves, but we want to integrate their proprietary user interfaces only. From one hand this approach allows us not to change



**Fig. 1. Database integration at Web interface level**

every integrated database structure dramatically (and thus established database utilization and administration technology – data update and insert). From the other hand, this approach allows to the end user to get access to the whole information picture on chemical substances that are stored in different databases. So called “virtual” database integration or in other words heterogeneous information system creation implies independence in evolution of separate subsystems and at the same time end user gets access to the whole information array on a particular chemical substance or material that is stored in databases of virtually united system (fig. 1). And that fact solves the main integration goal truly.

Taking into consideration current development conditions of Russian databases on physicochemical substances' properties the second integration approach – integration at interface level only – is more appropriate and quite perspective. It's worth mentioning that Web-interfaces have been developed for IMET RAS databases on physicochemical substance properties (“Crystal” database on acoustic-optical, electro-optical and non-linear optical substances and “Diagram” database on semiconductor systems phase diagrams). These Web-interfaces allow users to get remote access via Internet to data stored in these databases using any Web browser.

## Searching for Relevant Information in Integrated System

When integrating databases at Web-interfaces level it's required to provide facilities for browsing information contained in other databases. This information should be relevant to the data on some chemical system currently browsing by user. Let's consider this in the following example. User who browses information on Ga-As system from “Diagram” database should have an opportunity to get information for example on piezoelectric effect or non-linear optical properties of GaAs substance that contained in “Crystal” database. So it's obvious that, when designing distributed information system, it's required to provide search for relevant information that contained in other databases of distributed system. Thus, we hardly need to have some active data center that should know what information is contained in every integrated database. So some data store should exist that describes information contained in integrated database resources. In this manner, we come to the metabase concept – a special database that contains some reference information on integrated databases' contents (fig. 2). In our case, it is information on chemical systems and their properties. The amount of this metainformation should be enough to perform search for relevant information on systems and corresponding properties.

Let's try to formalize the problem in terms of set-theoretic approach. Hence, metabase should contain information on integrated databases ( $D$  set), information on chemical substances and systems ( $S$  set) and information on their properties ( $P$  set). To describe correlation between elements of  $D$ ,  $S$  and  $P$  sets let's define ternary relation called  $W$  on set  $U = D \times S \times P$ . Here  $U$  is a Cartesian product of  $D$ ,  $S$  and  $P$ . Membership of a  $(d, s, p)$  triplet to the  $W$  relation, where  $d \in D, s \in S, p \in P$ , can be interpreted in the following way: "Information on property  $p$  of chemical system  $s$  is contained in integrated database  $d$ ".

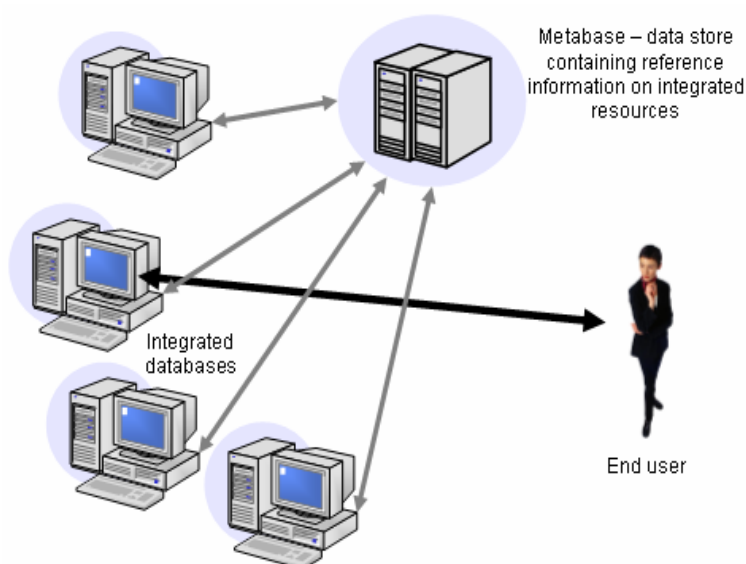


Fig. 2. Metabase concept

Having defined three basic sets it can be seen that search for information relevant to  $s$  system can be localized to determination of  $R$  relationship, that is a subset of Cartesian product  $S \times S$  (or in other words,  $R \subset S^2$ ). Thus, it can be stated about every pair  $(s_1, s_2) \in R$  that chemical system  $s_2$  is relevant to the system  $s_1$ . So all we need to solve the task of searching for relevant information in integrated databases is to determine somehow the  $R$  relation. It is significant to note that  $R$  relation can be created or complemented by means of either of two variants. The first variant is via using predefined rules by a computer. The second one is that experts in chemistry and materials science can be engaged to solve this task.

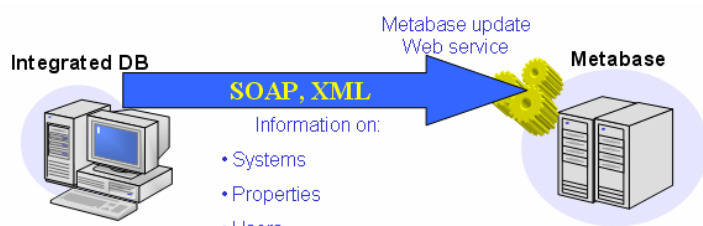
The second variant is quite clear – experts can form relationship  $R$  following some multicriterion rules affected by their expert assessments. So let's consider possible variants of automatic  $R$  relation generation. One of such variants can be like this one based on the following rules:

1. For any chemical systems  $s_1 \in S, s_2 \in S$ , that are composed from chemical elements  $e_{ij}$   $s_1 = \{e_{11}, e_{12}, \dots, e_{1n}\}, s_2 = \{e_{21}, e_{22}, \dots, e_{2m}\}$  it is true, that if  $s_1 \subseteq s_2$  (i.e. all chemical elements of system  $s_1$  are contained in system  $s_2$ ), then  $(s_1, s_2) \in R$ .
2.  $R$  relation is symmetric. In other words for any  $s_1 \in S, s_2 \in S$ , it is true, that if  $(s_1, s_2) \in R$ , then  $(s_2, s_1) \in R$  as well.

These two rules allow us to determine a set of chemical systems that are relevant to the given one. It should be noticed that this automatic  $R$  relation generation variant is just one of the simplest and most obvious variants of such rules, and in fact more complex mechanisms can be used to get  $R$  relation. For example, browsing information on a particular property of a compound in one of integrated databases (in fact, it is information defined by  $(d_1, s_1, p_1)$  triplet), we consider  $(d_2, s_2, p_2)$  triplet to be relevant information.  $(d_2, s_2, p_2)$  triplet characterizes information on some other property of a system from another integrated database. In this case, we have got more complex relevance relation like this  $R \subset (d_1, s_1, p_1) \times (d_2, s_2, p_2)$ , where  $d_1, d_2 \in D; s_1, s_2 \in S; p_1, p_2 \in P$ .

## Loading Information into Metabase

As it was stated above, Russian databases on materials science were developed in different organizations on various platforms and that fact makes integration significantly more complicated. Metabase should store reference data on integrated resources contents. It's obvious that in this situation it's required to use open network interconnection principles and standards that are supported on multiple platforms. If we consider present technology stack then it's quite clear that currently Web services are connection link between different platforms and heterogeneous environments. Web services are based on common standards such as SOAP (Simple Object Access Protocol) and XML (eXtensible Markup Language). Nowadays these technologies are capable of



**Fig. 3. Metabase update Web service**

providing reliable infrastructure for cross platform message exchange.

In that way reference information loading into metabase was implemented by means of metabase update Web service, so-called MUService (fig. 3). Let's consider metadata updating mechanisms in detail. System that is to be integrated with others should generate XML document that contains information on updates in that very system.

The layout format of this XML document is generally standardized for all integrated subsystems and it is strictly fixed by means of specially developed XML schema [1]. Thus, all subsystems being integrated should generate valid XML document that meets XML schema requirements to notify metabase of information changes that occurred in their state. After being generated, XML document is sent to the metabase update Web service for processing and metabase update. Interaction with this MUService is realized by means of SOAP protocol according to the Web service WSDL (Web Services Description Language) description [2]. In that way client databases report about updates to the metabase and so actual information on integrated resources contents appears in the metabase.

It's important to note that security issues were among primary concerns while designing and implementing the resulting system. In that way symmetric encryption mechanism was implemented with the aid of DES (Digital Encryption Standard). It guarantees secure metadata exchange with metabase update Web service. Additionally an option for data archiving was included into the system. A kind of zip-achieving was implemented that allows us to package data that are transmitted to the metabase server. This feature allows dramatically decreasing data volumes (taking into consideration high level of compression for XML documents) that are transmitted via public networks. Thus this feature lowers requirements to network bandwidth and it is very important and actual for Russia since high-speed Internet access is not available everywhere in the country. Compressing techniques enable us to decrease information volumes so that it becomes possible to use old-fashioned data modems on telephone wires to transmit data.

As it can be seen, metabase update Web service supports some rather complex additional data transformations (encryption and archiving) that require some extra coding. So to simplify the interaction process with the Web service special Web client was created. It was implemented as a COM object and thus it can be easily accessible from any environments that support Microsoft COM. Created Web client addresses issues connected with encryption and compression of information that is to be sent to the Web service. It controls all network interconnection aspects also. All this functionality just simplifies routine database attachment to the integrated system.

It should be mentioned that at present time only integrated database systems (as client systems to the metabase) could initiate data update process with metabase update Web service. This technique of course is not the only variant of interaction scheme. Thus it is planned to redesign metabase update mechanism so to enable metabase to inquire integrated resources on demand and thus to query information updates occurred.

It should be mentioned that after every metabase update session incremental population crawl is started on the metabase to update or to reindex relevant chemical systems list regarding information changes. This allows

metabase to maintain actual information on relevancy relation of chemical systems contained in integrated resources. Currently relevant system reindexing is performed by means of approach of two rules proposed in this paper earlier. If it is necessary, these rules can be easily modified. And the main advantage is that in that very case it is not necessary to redesign the whole system concept. All we have to do is just write a new piece of software to provide a new method of searching for relevant systems and replace the old module with the new one.

### Metabase Integration – How It Works

Let's consider the operating process of the integrated system from end user point of view. In a general sense, the integration of information resources of materials' science is in consolidation of available Web applications serving users of different materials' science databases. This consolidation is provided by means of specialized software but user should not be aware of it if possible. The software should be transparent in this sense.

When designing the integrated system special attention was paid to security system development. It should be mentioned that every developed information system has its own proprietary security facilities that protect the system and give permissions to access it. Security system of a particular information system is responsible for granting permissions to registered users of given system only. It's obvious that in the context of integrated security system authorized users should have permissions required getting access to the information in integrated resources within their privileges strictly.

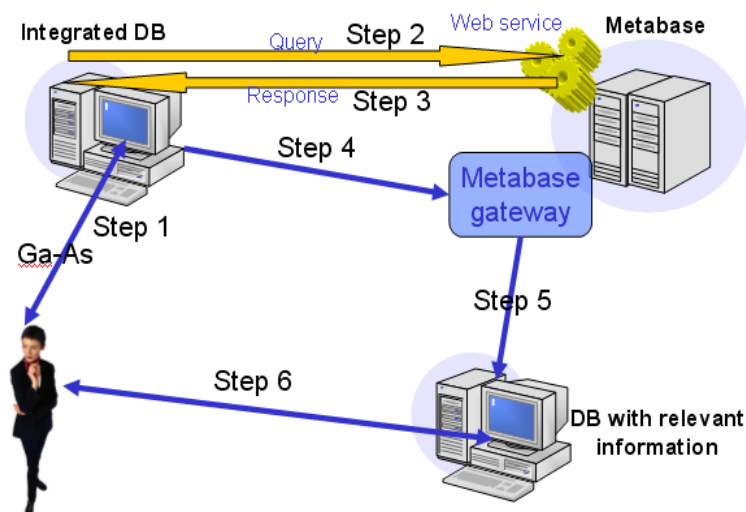


Fig. 4. Metabase integration – how it works.

For example, let's consider the possible user work session scenario. A user has been granted access to database on semiconductor system phase diagrams "Diagram" and currently he or she is browsing information on In-Sb system. Obviously the user should have an opportunity to get information on elastic constants of In-Sb system from "Crystal" database. But that user should not be granted privileges to observe information on chemical systems other than In-Sb since he or she is not a registered user of "Crystal" database. And vice versa, if the user is a registered user of "Crystal" database too then he or she will be granted full access to "Crystal" as integrated resource. From our point of view, the

described approach is an appropriate one and so it is used to design the distributed security system of integrated databases. It should be mentioned that user credentials of every integrated resource are also transmitted to the metabase via MUserService as well. It is done to organize distributed security system operation in cooperation with corresponding security systems of integrated resources. It should be emphasized that open user passwords are not transmitted to the metabase, instead of open passwords, password MD5 hashes are transmitted in fact. This substitution (MD5 hash instead of open password) allows integrated information system to authenticate active user and at the same time this technique excludes possibility of using open user password to login to the integrated system database. In other words, there is no place for vulnerabilities here. So even if this data are stolen integrated resources can not be compromised.

Let's assume that in one of integrated system user browses information on some particular chemical system. In other words, the user is in Web application of a particular information system (fig. 4). If it is necessary to get relevant information this Web application is capable to send a request to specially developed Web service [3] that serves users of integrated system. The request aim is to get information contained in integrated resources that is relevant to the currently browsed data. After the request Web service sends a response to the Web application in a form of XML document. It describes what relevant information on chemical systems and properties contained

in integrated resources. As it was mentioned, earlier data in XML format are properly understood on all major platforms. That information can be output to user for example by means of a XSL-transformation in form of HTML document (XML + XSL = HTML) containing hyperlinks to special gateway. The user can follow from one Web application to another to browse relevant information via this gateway only.

Imagine that the user clicks on a hyperlink to start browsing information from some other integrated system. First of all, when the user has clicked the hyperlink, he is forwarded to the special gateway. Actually it is a specialized Web application that runs on the metabase Web server. The gateway main purpose is to perform security-dispatching function in distributed system. According to the task stated it is responsible for user authentication and it also checks whether the user has required privileges to address the information requested. Let's imagine that authentication is successful and the user is eligible to address the data so the metabase security gateway performs redirection to a specialized entry point of desired Web application adding some additional information to create proper security context and a kind of digital signature. It should be stated that the entry point is a specialized page in target Web application that is to perform service functions for integrated system users. At this very page target Web application checks digital signature of the metabase security gateway and if everything is ok the page creates special security context for user with given access rights within target Web application. Finally, the user is automatically redirected to the page with the information required. In spite of redirection process apparent complexity, user transition from one Web application to another is absolutely transparent. Thus, end user can even not note that some complex processing has been done to perform redirection. So, it is an illusion created that having clicked on a hyperlink user just simply transferred from one information system to another.

---

## Conclusion

It's high time to draw a conclusion. The proposed database integration approach based on metabase was successfully applied at A.A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences (IMET RAS). "Crystal" and "Diagram" databases were the very first systems connected to the metabase integrated solution. This fact allows users of either information system to browse information from these databases. Now several words should be said about the project perspectives. First perspectives are connected with the resulting system extension due to addition of already developed Russian databases on materials science: IVTAN and MITHT databases. This integration will allow creating distributed database complex on electronics materials that has no analogs worldwide at present time. Besides numerical growth of integrated system there are plans for functional capabilities extension i.e. qualitative leap is planned. For example, it is projected to provide capability to perform complex distributed database queries that allow searching for substances that satisfy some defined complex criteria while information on criteria values is distributed over several databases. Consequently, to successfully perform such query it's required that metabase information system has an opportunity to query distributed databases impersonating acting user who initiates the initial complex query. After that the metabase should gather information from several sources, process it and output to the end user. Integration at that level undoubtedly will expand distributed information resources capabilities significantly.

---

## Bibliography

- [1] XML-schema that standardized XML document format for metabase update Web service is available at <http://meta.imet-db.ru/MUService.xsd>
- [2] WSDL-contract that defines methods that can be utilized to interact with metabase update Web service is available at <http://meta.imet-db.ru/MUService/MUService.asmx?wsdl>
- [3] WSDL-contract that defines methods that can be utilized to interact with Web service that serves integrated resources is available at <http://meta.imet-db.ru/Service/Service.asmx?wsdl>

---

## Authors' Information

**Valery Kornyshko** – MITHT, Head of IT department; 119571, pr. Vernadskogo, 86, Moscow, Russia; e-mail: [inftech@mitht.ru](mailto:inftech@mitht.ru)

**Victor Dudarev** – MITHT, junior member of teaching staff of IT department; 119571, pr. Vernadskogo, 86, Moscow, Russia; e-mail: [vic@osg.ru](mailto:vic@osg.ru)



## ANALYZING THE DATA IN OLAP DATA CUBES\*

Galina Bogdanova, Tsvetanka Georgieva

**Abstract:** OLAP applications provide a possibility to data analysis over large collections of historical data in the data warehouses, supporting the decision-making process. This paper presents an application that creates a data cube and demonstrates the effectiveness of the applying the OLAP operations when it necessary to analyze the data and obtain the valuable information from the data. It allows the analysis of factual data that is daily downloads of folklore materials, according to dimensions of interest.

**Keywords:** data cube, online analytical processing, multidimensional expressions

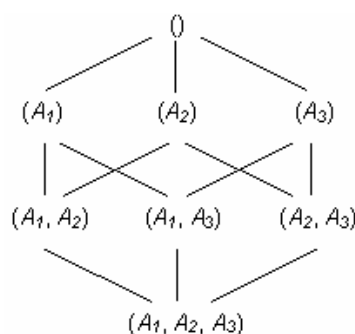
### 1. Introduction

Decision-support functions in a data warehouse, such as online analytical processing (OLAP), involve hundreds of complex aggregate queries over large volumes of data. It is not feasible to compute these queries by scanning the data sets each time [9]. The data cubes are structures designed to provide quick access to the data in data warehouses. The cube definition is determined from the requirements, which the users analyzing the data have and it is based on the choice of the schema representing dimensional model of the data that consists of one fact table and several dimension tables.

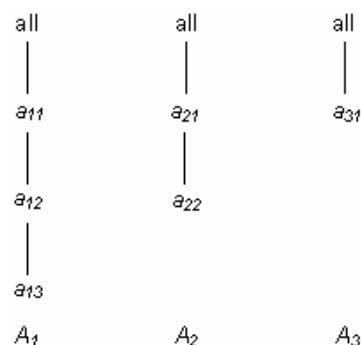
Some requirements of decision-support systems (DSS) are considered in [4]. Several common considerations of designing data cubes are examined in [5]. In [11] is presented an OLAP server supporting dimension updates and view maintenance under these updates. The advantages of the aggregation ranking queries in OLAP data cubes are described in [6]. In this paper are applied the OLAP operations to analyze the data in a WEB based client/server system containing archival fund with folklore materials of the Folklore Institute at BAS. It represents the data cube creation and the dimension hierarchies. Part of present paper is reported in [2].

The rest of the paper is organized as follows. Section 2 reviews the concepts of the data cubes, the lattice corresponding to a data cube and the OLAP operations. Section 3 presents the data cube creation and applying the OLAP operations by using the language MDX (*Multidimensional Expressions*). Section 4 gives the conclusion of this paper.

### 2. Data Cubes and OLAP Operations



(a) Data cube lattice



(b) Dimension hierarchy lattices

Fig. 1 Example lattices

\* Supported partially by the Bulgarian National Science Fund under Grant MM-1405/2004

Data cubes are popular in OLAP because they provide an intuitive way for data analysts to navigate various levels of summary information in the database [9]. In a data cube, attributes are categorized into *dimension attributes*, on which grouping may be performed, and *measures*, which are the results of aggregate functions. From a data cube with  $n$  dimension attributes can be obtained  $2^n$  cube views. For example, a data cube with dimensions  $A_1, A_2, A_3$  is shown in figure 1(a) as a lattice structure.

The dimensions often are organized into dimension hierarchies, which can also be represented by a lattice. For example, figure 1(b) shows the lattice for the dimension hierarchies where  $a_{ij}$  is  $j$ -th level in the hierarchy of the dimension  $A_i$ . The top element of each lattice is "all", meaning no grouping by that dimension.

It can be constructed a lattice representing the set of views that can be obtained by grouping on each combination of elements from the set of the dimension hierarchies. Figure 2 shows the lattice combining the data cube lattice of figure 1(a) with the dimension hierarchy lattices of figure 1(b).

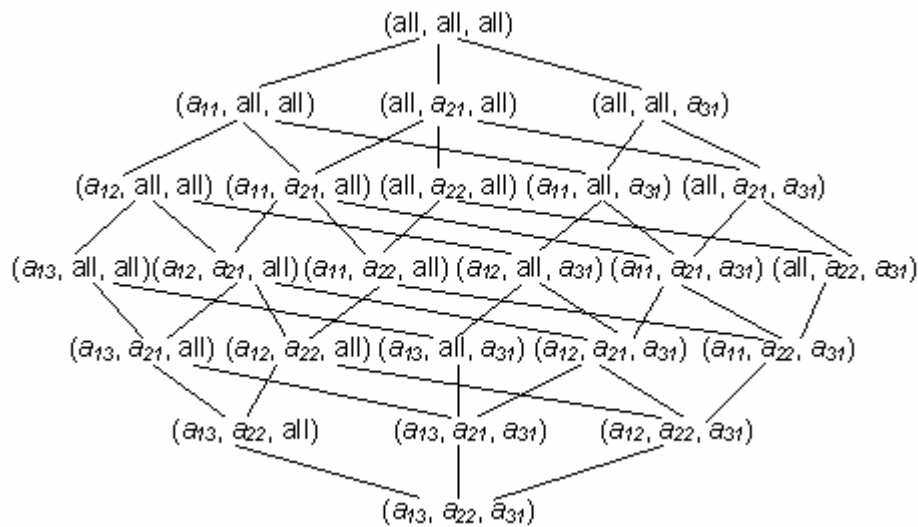


Fig. 2 Combined lattice

OLAP includes a set of operations for manipulation of the dimensional data organized in multiple levels of abstraction. The basic OLAP operations are roll-up (increasing the level of aggregation); drill-down (decreasing the level of aggregation); slice-and-dice (selection and projection); pivot (re-orienting the multidimensional view of data) [3].

### 3. Creation of the Data Cube FolkloreCube and Analyzing the Data by Applying the OLAP Operations

The investigated archive keeps detailed information of the documents and materials, which can be downloaded by the users and contain audio, video and text information.

#### 3.1. The Relational Database FolkloreDB

The OLTP (online transaction processing) database FolkloreDB is created in accordance to the classification schema described in [7]. This database is realized by using the client/server relational database management system (RDBMS) Microsoft SQL Server [8, 10] and consists of the tables shown in figure 3.

#### 3.2. The Data Warehouse Database FolkloreDB\_DW

The database FolkloreDB\_DW in data warehouse is designed by using the dimensional model represented by the star schema. This database consists of one fact table Downloads\_fact and four dimension tables Documents\_DW, Links\_DW, Users\_DW and Dates (fig. 4).

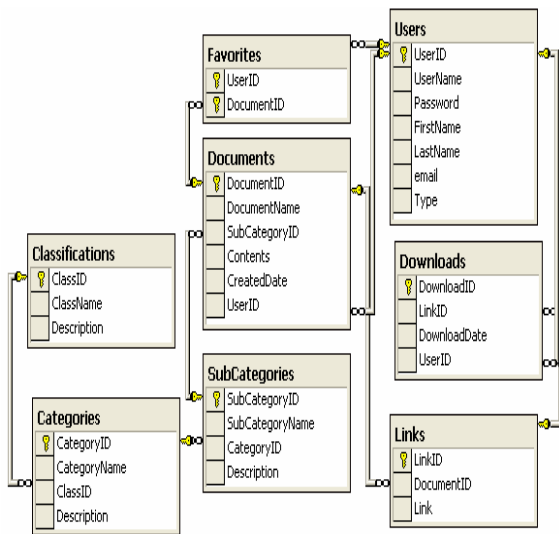


Fig. 3 The structure of the database FolkloreDB

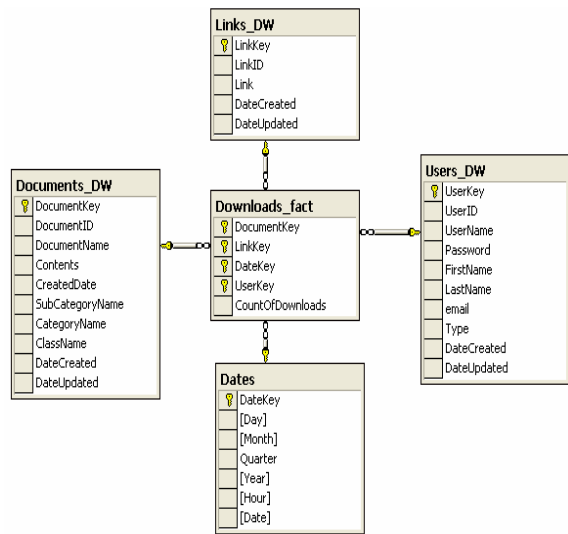


Fig. 4 The structure of the database FolkloreDB\_DW in the data warehouse

### 3.3. Creation of the Data Cube FolkloreCube

The data cube FolkloreCube is created in correspondence with the star schema of the dimensional model of the database FolkloreDB\_DW. The represented statement defines the data cube structure.

```
CREATE CUBE FolkloreCube(
    DIMENSION Document,
        LEVEL [All Documents] TYPE ALL,
        LEVEL [Class name],
        LEVEL [Category name],
        LEVEL [SubCategory name],
        LEVEL [Document name],
    DIMENSION [Link],
        LEVEL [All Links] TYPE ALL,
        LEVEL [Link name],
    DIMENSION [User],
        LEVEL [All Users] TYPE ALL,
        LEVEL [Type],
        LEVEL [User name],
    DIMENSION [Time] TYPE TIME,
        LEVEL [All Time] TYPE ALL,
        LEVEL [Year] TYPE YEAR,
        LEVEL [Quarter] TYPE QUARTER,
        LEVEL [Month] TYPE MONTH,
        LEVEL [Day] TYPE DAY,
        LEVEL [Hour] TYPE HOUR,
    MEASURE [Count of downloads]
        Function Sum)
```

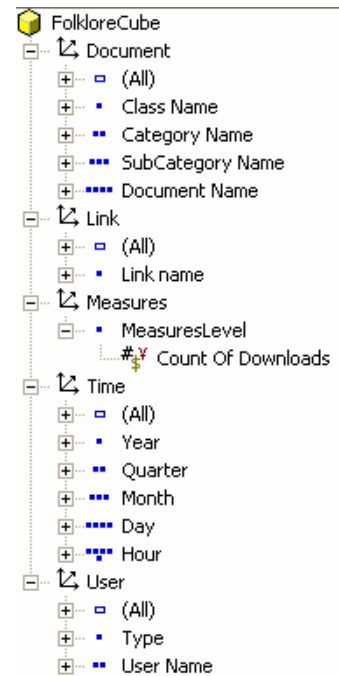


Fig. 5 The structure of the data cube FolkloreCube

It is executed from the Visual Basic project [1, 12] and creates the local cube (.cub) by using Microsoft ActiveX Data Objects and ActiveX Data Objects Multidimensional (ADO MD).

This statement adds data into already defined cube:

```
INSERT INTO FolkloreCube
(Document.[Class name], Document.[Category name],
 Document.[Subcategory name], Document.[Document name],
 Link.[Link name], User.[Type], User.[User name], Time.[Year],
 Time.[Quarter], Time.[Month], Time.[Day], Time.[Hour],
 Measures.[Count of downloads])
SELECT d.ClassName, d.CategoryName, d.SubCategoryName,
      d.DocumentName, l.Link, u.Type, u.UserName, t.Year,
      t.Quarter, t.Month, t.Day, t.Hour, f.CountOfDownloads
FROM Documents_DW d, Downloads_fact f, Links_DW l, Users_DW u,
      Dates t
WHERE d.DocumentKey = f.DocumentKey AND f.LinkKey = l.LinkKey
      AND u.UserKey = f.UserKey AND t.DateKey = f.DateKey
```

The dimension hierarchies of the data cube FolkloreCube are shown in figure 5.

### 3.4. MDX Queries

MDX is Microsoft OLAP query language [13, 14, 15]. MDX queries are applied to the data cube FolkloreCube providing the dimensional view of summarized data.

#### 3.4.1. Top 5 documents from which are downloaded materials (fig. 6);

The screenshot shows a web-based OLAP client interface with a pivot table. The table has three columns: 'Users', 'Period of time', and 'Count of downloads'. The 'Users' column lists five document categories: 'Годож', 'Курбан', 'Кръщение', 'Коледна приказка - текст', and 'Имен ден'. The 'Count of downloads' column shows the corresponding values: 82, 10, 7, 5, and 4. The interface also includes a sidebar with various OLAP operations like 'Top 5 documents from which are downloaded materials', 'Count of the documents by categories', etc.

Users	Period of time	Count of downloads
Годож		82
Курбан		10
Кръщение		7
Коледна приказка - текст		5
Имен ден		4

Fig. 6 Exemplary result from applying OLAP operations in FolkloreCube

The execution of the following MDX query provides the result represented in figure 6:

```
SELECT {Measures.[Count of downloads]} ON COLUMNS,
      TOPCOUNT(Document.[Document name].Members, 5,
      Measures.[Count of downloads]) ON ROWS
FROM FolkloreCube
```

3.4.2. The documents from which are downloaded the most of materials from the top 10 users (fig. 7);

Users	Period of time	Count of downloaded materials
потребител1	Годож	79
потребител1	Курбан	9
потребител1	Кръмене	5
потребител1	Коледна приказка - текст	5
потребител1	Изпращане на войник	3
потребител2	Годож	2
потребител2	Имен ден	1
потребител2	Кръмене	1
потребител2	Курбан	1
потребител2	Семейство	1
потребител3	Годож	1
потребител3	Кръмене	1
потребител3	Сватба	1
потребител3	Семейство	1
потребител3	Изпращане на войник	0

Fig. 7 Example for 3-dimensional view of the data

To obtain the result show in figure 7 the represented application performs the MDX query:

```
WITH SET Top10Users AS 'TOPCOUNT(User.[User name].Members, 10,
    Measures.[Count of downloads])'
MEMBER Measures.[Count of downloaded materials] AS
    'CoalesceEmpty(Measures.[Count of downloads], 0)'
SELECT {Measures.[Count of downloaded materials]} ON COLUMNS,
    {GENERATE(Top10Users, CROSSJOIN({User.CURRENTMEMBER},
        TOPCOUNT(Document.[Document name].Members, 5,
            Measures.[Count of downloads])))} ON ROWS
FROM FolkloreCube
```

3.4.3. Count of the materials downloaded from documents by the hours of the chosen date and the difference with the previous hour (fig. 8).

Period of time	Count of downloads	Difference with previous hour
For chosen day	40	
00	1	0
01	1	0
02	1	0
04	1	0
05	1	0
07	1	0
10	1	0
14	13	12
15	14	1
16	3	-11
17	2	-1
18	1	-1

Fig. 8 Analyzing the data in FolkloreCube for given period of time

```
WITH SET M AS '{Time.[2004].[Q4].[11].[27]}'
MEMBER Time.[For chosen day] AS 'M.Item(0)'
MEMBER Measures.[Difference with previous hour] AS
    '(Measures.[Count of downloads], Time.CURRENTMEMBER) -
    (Measures.[Count of downloads],
```

```

        Time.CURRENTMEMBER.PrevMember) '
SELECT {Measures.[Count of downloads],
        Measures.[Difference with previous hour]} ON COLUMNS,
        {Time.[For chosen day],
        GENERATE(M, {Time.CURRENTMEMBER.children})} ON ROWS
FROM FolkloreCube

```

---

#### 4. Conclusion

With rapid developments of data warehouses and OLAP technologies and with enormous amount of data stored in databases in result of daily activities of the organizations it is increasingly important to develop the database applications that convert huge volumes of data into meaningful information. This assists in the decision-making process in different area by providing feedback on past actions of the users and helped to guide future decisions. Represented application gives different views of the data collected in a WEB based client/server system that contains archival fund with folklore materials.

---

#### Bibliography

- [1] Bekuit, B., VB.NET: A Beginner's Guide, AlexSoft, 2002, pages 330.
- [2] Bogdanova, G., Tsv. Georgieva, Applying the OLAP Operations to Analyzing the Data in a WEB based Client/Server System Containing Archival Fund with Folklore Materials, National Workshop on Coding Theory, Bankya, 9-12.12.2004.
- [3] Garcia-Molina, H., J. D. Ullman, J. Widom, Database Systems: The Complete Book, Williams, 2002, pages 1083.
- [4] Georgieva, Tsv., Data Warehousing and OLAP Technology, In Proc. of the International Conference on Computer Systems and Information Technologies, Veliko Tarnovo, 3-5.10.2003, pages 58-65 (in bulgarian).
- [5] Georgieva, Tsv., Designing OLAP Data Cubes, In Proc. of the International Conference on Computer Systems and Information Technologies, Veliko Tarnovo, 3-5.10.2003, pages 106-112 (in bulgarian).
- [6] Li, H., H. Yu, D. Agrawal, A. Abbadi, Ranking Aggregates, Technical Report 2004-07, Department of Computer Science, University of California, Santa Barbara, 2004, pages 14.
- [7] Mateeva, V., I. Stanoeva, Classification Scheme of the Typological Catalogue in the Folklore Institute, Bulgarian Folklore, v. 2-3, 2001, pages 96-109 (in bulgarian).
- [8] Microsoft Corporation, MCSE Training: Microsoft SQL Server 2000 Database Design and Implementation, SoftPress Ltd., 2001, pages 800.
- [9] Mumick, I., D. Quass, B. Mumick, Maintenance of Data Cubes and Summary Tables in a Warehouse, In Proc. ACM SIGMOD Conf. on Management of Data, Tuscon, Arizona, 1997, pages 100-111.
- [10] Soukup, R., K. Delaney, Inside Microsoft SQL Server 7.0, SoftPress Ltd., 2000, pages 952.
- [11] Vaisman, A., A. Mendelzon, W. Ruaro, S. G. Cymerman, Supporting Dimension Updates, In Proc. of the 14<sup>th</sup> International Conference on Advanced Information Systems Engineering, 2002, pages 67-82.
- [12] Wang, W., Visual Basic 6: A Beginner's Guide, AlexSoft, 2002, pages 562.
- [13] <http://www.georgehernandez.com/xDatabases/MD/MDX.htm>
- [14] <http://www.microsoft.com/data/oledb/olap>
- [15] <http://www.microsoft.com/sql>

---

#### Authors' Information

**Galina Bogdanova** – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Veliko Tarnovo, P.O.Box: 323, e-mail: [galina@moi.math.bas.bg](mailto:galina@moi.math.bas.bg)

**Tsvetanka Georgieva** – University of V. Tarnovo "St. St. Cyril and Methodius", Department of Information Technologies; e-mail: [cv.georgieva@uni-vt.bg](mailto:cv.georgieva@uni-vt.bg)

---

---

## Knowledge Engineering

---

---

### CLASSIFICATION OF BIOMEDICAL SIGNALS USING THE DYNAMICS OF THE FALSE NEAREST NEIGHBOURS (DFNN) ALGORITHM<sup>1</sup>

Charles Newton Price, Renato J. de Sobral Cintra,  
David T. Westwick, Martin P. Mintchev

**Abstract:** *Accurate and efficient analysis of biomedical signals can be facilitated by proper identification based on their dominant dynamic characteristics (deterministic, chaotic or random). Specific analysis techniques exist to study the dynamics of each of these three categories of signals. However, comprehensive and yet adequately simple screening tools to appropriately classify an unknown incoming biomedical signal are still lacking. This study is aimed at presenting an efficient and simple method to classify model signals into the three categories of deterministic, random or chaotic, using the dynamics of the False Nearest Neighbours (DFNN) algorithm, and then to utilize the developed classification method to assess how some specific biomedical signals position with respect to these categories. Model deterministic, chaotic and random signals were subjected to state space decomposition, followed by specific wavelet and statistical analysis aiming at deriving a comprehensive plot representing the three signal categories in clearly defined clusters. Previously recorded electrogastrographic (EGG) signals subjected to controlled, surgically-invoked uncoupling were submitted to the proposed algorithm, and were classified as chaotic. Although computationally intensive, the developed methodology was found to be extremely useful and convenient to use.*

**Keywords:** *Biomedical signals, classification, chaos, multivariate signal analysis, electrogastrography, gastric electrical uncoupling*

---

#### 1. Introduction

Efficient accumulation of accurate knowledge from a wide variety of biomedical phenomena can be obtained from studying and analyzing their dynamics. This dynamics can be assessed by various sensors which monitor, measure, and transform biomedical phenomena into electrical signals that can be analyzed using contemporary electronics and signal processing techniques [1]. Generally, biomedical signals can be to an extent deterministic, random or chaotic [1, 2]. Deterministic signals have the characteristic of predictability, meaning that any future course of the signal could be predicted using some linear analysis tools [1]. Random signals are non-deterministic, in the sense that individual data points of the signal may occur in any order [1], limiting determining the predictability of the future course of the signal to purely stochastic analytical tools. Chaotic signals can be viewed as a connecting mesh between deterministic and random signals, exhibiting behaviour that is slightly predictable, non-periodic, and highly sensitive to initial conditions [2].

Observing that there are three general types of biomedical signals that can be encountered, accurate and efficient study and analysis of these signals can be facilitated by their proper identification as deterministic, random or chaotic, given that specific analysis techniques exist for each type of signals [1]. However, comprehensive and yet adequately simple screening tools to appropriately classify an unknown incoming biomedical signal with respect to these three categories are still lacking.

---

<sup>1</sup> This study was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the Gastrointestinal Motility Laboratory (University of Alberta Hospitals) in Edmonton, Alberta, Canada

Recent paper by Gautama et al. [3] presents a method to classify an unknown incoming biomedical signal. The method provides an interpretation of a signal's deterministic and/or stochastic nature in terms of its predictability. Furthermore, it assesses the signal's linear or non-linear nature using surrogate data methods [3]. The result of this study provides a tool to measure the amount of determinism and randomness in a biomedical signal, useful for detecting a change in health conditions from monitored biomedical signals. However, this method can be seen as an analysis technique that can be applied to a signal once it is classified as deterministic, chaotic or random, rather than as a signal classifier.

The aim of the present work was to develop an efficient and simple method to classify biomedical signals into three categories (deterministic, chaotic or random), using a novel chaos analysis technique which we called the Dynamics of the False Nearest Neighbors (DFNN) algorithm. The proposed method extends the previously developed False Nearest Neighbors (FNN) algorithm [2, 4], to include dynamic FNN characteristics.

---

## 2. Methods

---

Understanding the suggested technique requires an introduction to multivariate signal analysis using state space representation, including time delay and embedding dimension calculations [2, 4, 5].

### 2.1. State Space Signal Representation

Biomedical signals are usually observed in one-dimensional form, and are represented discretely in the form of a time-domain vector,  $\mathbf{s}(n)$ . It can be inferred that the one-dimensional time-domain vector,  $\mathbf{s}(n)$ , is a projection of the signal generator source, represented by an unknown but underlying multidimensional dynamic state vector  $\mathbf{x}(n)$  [2]. The multidimensional dynamic state vector is composed of an unknown number of variables, represented through its dimension  $d$  [2, 6]. In these notations  $n$  denotes the current moment in the sampled time-domain.

The transition from a sampled one-dimensional time-domain signal  $\mathbf{s}(n)$  to the corresponding sampled  $d$ -dimensional state space requires the application of Takens Theorem [6]. Takens Theorem represents a technique to reconstruct an approximation of the unknown dynamic state vector  $\mathbf{x}(n)$  in  $d$ -dimensional state space by lagging and embedding the observed time series  $\mathbf{s}(n)$ . This reconstructed approximation is the state vector  $\mathbf{y}(n) = [s(n), s(n+T), s(n+2T), \dots, s(n+T(d-1))]$ , composed of time-delayed samples of  $\mathbf{s}(n)$ , where  $T$  is the time delay and  $d$  is the embedding dimension of the system. The accurate calculation of  $d$  and  $T$  guarantees through the Embedding Theorem [2], that the sequential order of the reconstructed state vector  $\mathbf{y}(n) \rightarrow \mathbf{y}(n+1)$  is topologically equivalent to the generator state vector  $\mathbf{x}(n) \rightarrow \mathbf{x}(n+1)$ , allowing  $\mathbf{y}(n)$  to represent without ambiguity the actual source of the observed multidimensional dynamic vector  $\mathbf{x}(n)$  [2].

Each state space coordinate  $[s(n), s(n+T), s(n+2T), \dots, s(n+T(d-1))]$  constituting a component of  $\mathbf{y}(n)$  defines a point in the state space. As time progresses, the dynamic trajectory of each point in time forms what is called an orbit. An orbit is mathematically defined as the numerical trajectory resulting from the solution of the system [2]. Each orbit constituting  $\mathbf{y}(n)$  is presumed to come from an autonomous set of equations, and therefore, according to the Uniqueness Theorem [2], the trajectory of any orbit is unique and should not overlap with itself. The time delay  $T$  is an integer multiple of the sampling interval of the signal  $\mathbf{s}(n)$  guaranteeing the extraction of maximal amount of information from the system [2]. The embedding dimension  $d$  is the minimal state space dimension required to unfold the main orbit of  $\mathbf{x}(n)$  [2]. The main orbit of  $\mathbf{x}(n)$ , known as the attractor, represents the set of points in state space visited by the other orbits of the system long after transients have died out [2].

### 2.2. Time Delay Calculation

The choice of an accurate time delay  $T$  guarantees that the time-delayed state space coordinates forming  $\mathbf{y}(n)$  are independent from each other [2]. Choosing too small of a value for  $T$  clusters the data in state space, while choosing too large of a value for  $T$  causes the disappearance of the relationships between the points in the attractor [7, 8]. The independence between two coordinates of  $\mathbf{y}(n)$  can be assessed using the mutual information (MI) function [2]. The MI between two  $\mathbf{y}(n)$  coordinates, e.g.,  $s(n)$  and  $s(n+T)$ , is measured in bits by:

$$MI = \log_2 \left\{ \frac{P[s(n), s(n+T)]}{P[s(n)]P[s(n+T)]} \right\}, \quad (1)$$



where  $P[s(n), s(n+T)]$  is the joint probability density function (JPDF) of  $s(n)$  and  $s(n+T)$ . The average mutual information (AMI) of the JPDFs of all coordinates is calculated by:

$$AMI(T) = \sum_{s(n), s(n+T)} P[s(n), s(n+T)] \log_2 \left\{ \frac{P[s(n), s(n+T)]}{P[s(n)]P[s(n+T)]} \right\}. \quad (2)$$

The first minimum of the AMI function provides the optimal time delay  $T$ , and assures the independence between the coordinates of the multidimensional vector  $\mathbf{y}(n)$  [2, 7, 8].

### **2.3. Embedding Dimension Calculation**

The signal reconstruction in state space requires a dimension that will guarantee no overlap of the trajectory of the orbit constituting  $\mathbf{y}(n)$ . This optimal dimension is obtained after calculating the percentage of False Nearest Neighbours (FNN) between points in state space. FNNs are calculated using reconstructed state space vectors  $\mathbf{y}(n)$  at different embedding dimensions but a constant time-delay [9]. It is accepted that when the FNN percentage drops to zero, the minimum required dimension to unfold the system into its original state around its attractor is reached, which also guarantees that the orbit is unique [2, 9]. The calculation of the FNNs requires the measurement of a distance  $R_d$ , defined as the radius between neighbouring vectors in consecutive dimensions.

This procedure is referred to as the FNN algorithm [2, 9]. The square of the Euclidian distance representing  $R_d$  as seen in dimension  $d$  is:

$$R_d(n)^2 = \sum_{m=1}^d [s(n+T(m-1)) - s^{NN}(n+T(m-1))]^2, \quad (3)$$

where  $n$  is the current index of the discrete signal (in this case  $s(n)$ ) and  $s^{NN}$  is the nearest neighbour (NN) of  $s(n)$ .

The square of the Euclidian distance in dimension  $d+1$  becomes:

$$R_{d+1}(n)^2 = \sum_{m=1}^{d+1} [s(n+T(m-1)) - s^{NN}(n+T(m-1))]^2 = R_d(n)^2 + (s(n+dT) - s^{NN}(n+dT))^2 \quad (4)$$

The change in distance between the points at dimensions  $d$  and  $d+1$  is:

$$\sqrt{\frac{R_{d+1}^2(n) - R_d(n)^2}{R_d(n)^2}} = \frac{|s(n+dT) - s^{NN}(n+dT)|}{R_d(n)}. \quad (5)$$

Determining the existence of a false nearest neighbour depends on how the distance between state space vectors behaves as the calculations progress in consecutive dimensions. If the distance increases significantly with the increment of the embedding dimension, then the vectors are false neighbours, and their closeness results from the reconstruction dynamics of the system, not from its underlying dynamics [2, 9]. If the distance is restricted within a certain threshold level close to the state space points, then the state space points are real neighbours resulting from the dynamics of the system. The embedding dimension that adequately represents the system is the dimension that eliminates most of the false neighbours, leaving a system whose trajectories are positioned in state space due to their underlying dynamics, not to their reconstruction dynamics. Figure 1 shows an example of the results of the FNN algorithm applied to model deterministic, chaotic and random signals.

### **2.4. Dynamics of FNN (DFNN) Algorithm**

The FNN algorithm is utilized to determine the minimal embedding dimension required to completely reconstruct in state space the source  $\mathbf{x}(n)$  of a one-dimensional time series  $\mathbf{s}(n)$  [2, 9]. Theoretically, the minimal embedded dimension is obtained when the percentage of FNN at a given dimension reaches zero. In practice, not all signals tested through the FNN algorithm reach zero percent FNNs. Two main factors are responsible for this fact. The first is that the larger the embedding dimension of the system, the larger the number of signal samples required

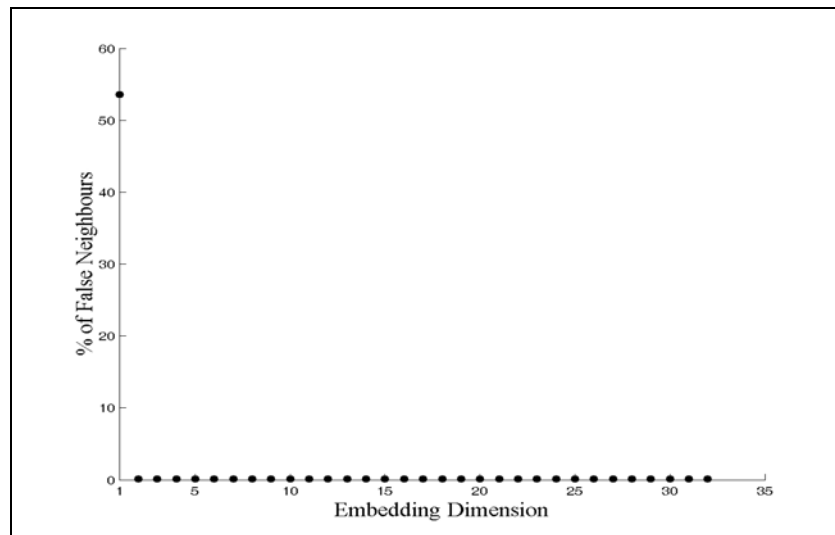
for the FNN algorithm [2]. The minimum number of samples required for a given embedding dimension is given by the following equation:

$$m_d = \sqrt{2} \times (e)^d, \quad (6)$$

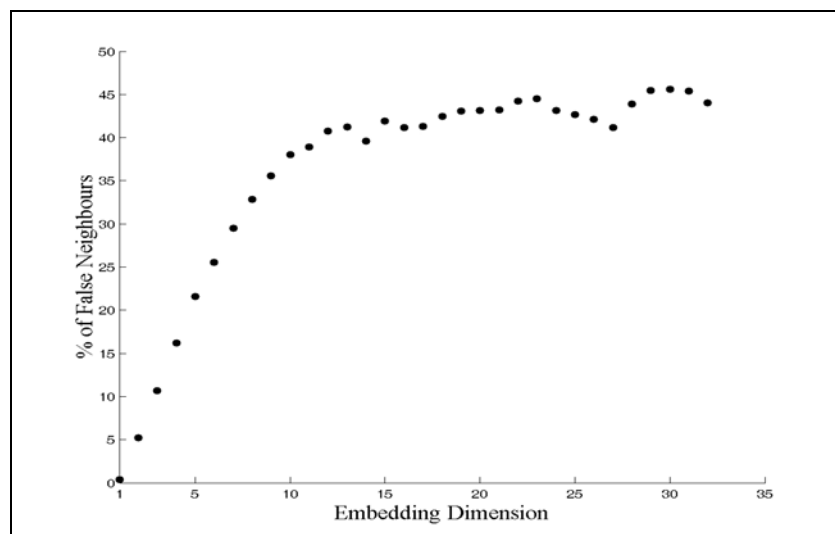
where  $d$  is the number of embedding dimensions. The second factor is that an established pattern, or attractor, underlying the dynamics of the system, simply does not exist, as it is the case with white noise [5], for which the FNN algorithm does not reach zero percent FNNs. Therefore, an FNN algorithm failing to converge to zero percent FNN indicates a signal which dimension is too high for the number of samples available.

The important benefit of the proposed DFNN algorithm is that it analyzes the reconstruction dynamics of the signals submitted to the FNN algorithm. Thus, in contrast to the well-established FNN algorithm [9], the DFNN approach does not aim at finding the optimal embedding dimension of the processed signals, but focuses on the processing of the curve representing the FNN dynamics as a function of the embedded dimensions (see Figure 1).

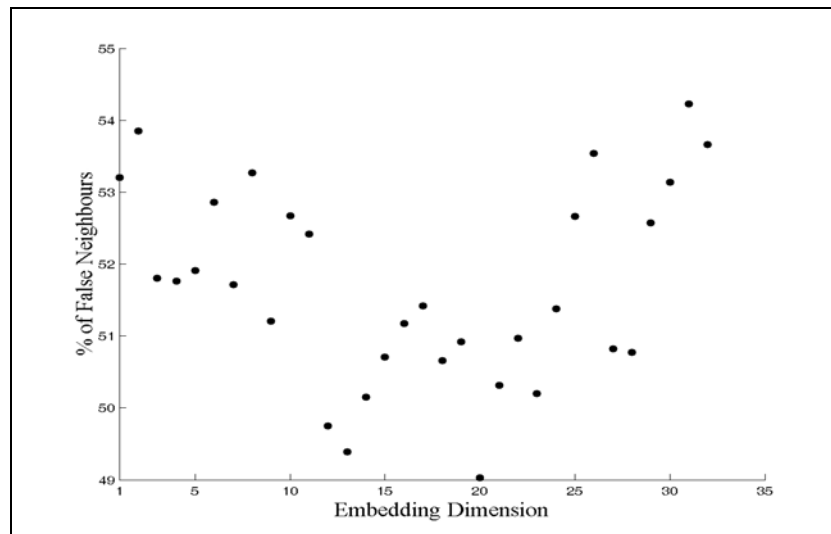
Figure 1 – Sample FNN dynamics of models of (a) deterministic, (b) chaotic, and (c) random signals.



(a) deterministic



(b) chaotic



(c) random signals

The independence of the coordinates of the reconstructed state space vector  $\mathbf{y}(n)$  is guaranteed by an accurate choice of time delay  $T$  [2, 7, 8]. According to the Uniqueness Theorem [2], each reconstruction is unique, therefore the use of an inaccurate embedding dimension when the signal is reconstructed still guarantees a unique representation of the signal, even though it would not represent properly its underlying dynamics until the correct embedding dimension is calculated with the help of the FNN algorithm. What this implies is that the percent of FNNs incrementally calculated at different embedding dimensions in the algorithm is a unique property of the signal under consideration, and therefore, the dynamics of these percent FNNs could also be regarded as a way to represent the signal.

The proposed DFNN algorithm uses a wavelet based pattern recognition technique to classify sampled signals as deterministic, chaotic, or random. The technique is based on analyzing the reconstruction dynamics of the FNNs at consecutively increasing embedding dimensions, ranging from 1 to 32 with the help of wavelet decomposition [10]. The limit of 32 was established due to the fact that 32 is an adequate number of points for a meaningful statistical analysis, because for dyadic wavelet analysis [11] a number that is a power of two is needed, and because it has been shown that the Rossler chaotic system can have an embedding dimension of 25 [12]. It is important to clarify that our aim was not to find the optimal embedding dimension of each signal, but to analyze their reconstruction dynamics as they were submitted to the FNN algorithm.

## **2.5. Wavelet Decomposition**

Wavelet decomposition analysis can be utilized to quantify the shape of the data points extracted from the FNN algorithm [13]. Wavelet analysis coefficients can indicate the resemblance between the shape of a wavelet and a signal. If the resemblance is high, the signal energy is concentrated in few wavelet coefficients. Otherwise, the energy content of the signal is spread throughout these coefficients [14]. Therefore, the aim to quantify the shape of the data points extracted from the FNN algorithm requires (i) finding the wavelet that matches best the wave-shape of the FNN dynamics [14], and (ii) calculating the corresponding wavelet analysis coefficients.

The proposed DFNN algorithm attempts to classify sampled signals as deterministic, chaotic, or random. An example wave-shape for each signal can be seen in Figure 1. Notice that the shape of the FNN dynamics for each type of signal is different, and that in order to make a comparison one signal type needs to be chosen to be a reference. The deterministic sampled signal pattern was chosen as a reference for determining the wavelet due to its simple shape. The Haar system wavelet was selected, since it matched best the pulse-like wave-shape of deterministic FNN dynamics (compare Figure 1a to Figure 2). Therefore, we hypothesized that through the analysis of the wavelet coefficients, a distinction could be made between deterministic, chaotic and random signal patterns using the FNN dynamics associated with a particular signal.

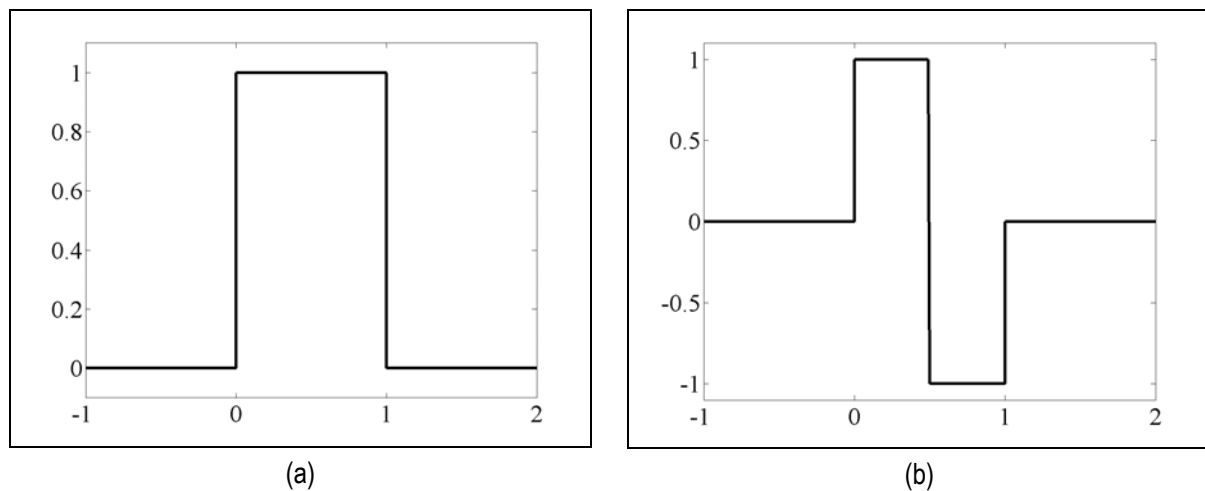


Figure 2 – (a) Haar scaling function and (b) wavelet.

## 2.6. Testing the DFNN Algorithm

### 2.6.1. Categories of Signal Models

In order to test the validity of the proposed DFNN algorithm, signal models with known characteristics were utilized, classified in the existing literature as deterministic, chaotic and random signals [1, 2].

All model signals were sampled frequently enough to comply with the Nyquist Theorem, guaranteeing sufficient digital samples to represent each signal [15]. The length of each signal was 6000 points. The deterministic group of signals included sine, rectangular, triangular, square, sawtooth, quadratic and Dirichlet functions. The models of chaotic signals included the Mackey-Glass Map (MGM), Henon, Ikeda and logistic maps, as well as the Lorenz, and quadratic systems. The random signal realizations were based on the following statistical distributions: Rayleigh, exponential, beta,  $\chi^2$  (chi-squared), gamma, impulsive, normal, and uniform.

### 2.6.2. Testing Protocol

The three categories of model signals (deterministic, chaotic and random) were submitted to the FNN algorithm, where the percent FNN was calculated and recorded for each embedding dimension up to 32. The resulting 32-point representation of each signal (see samples in Figure 1), were submitted to wavelet analysis decomposition in order to evaluate its corresponding wavelet coefficients. The wavelet coefficients were calculated for each model signal in each corresponding category (deterministic, chaotic and random) and were statistically analyzed by calculating the sample mean and standard deviation. The sample mean and standard deviation were considered distinguishable features to be submitted to the Fuzzy C-means clustering algorithm [16]. The Fuzzy C-means algorithm groups these distinguishable features into clusters. Each cluster provides a centroid, representative of each model signal [16]. This centroid is an important feature corresponding to each cluster that was pivotal for categorizing each signal.

### 2.6.3. Assessing the Robustness of the DFNN Algorithm

The limited amount of model signals used to test the DFNN algorithm could affect the statistical significance of the results. Therefore, two additional tests were designed to further strengthen the robustness of the DFNN algorithm. The first test involved shuffling the data points of deterministic and chaotic signals to transform them into random signals. These shuffled surrogate signals were then submitted to the DFNN algorithm, and categorized according to their position relative to their nearest cluster using the similarity measure (SM) [17].

The chosen SM is based on the Euclidian metric and is represented by a number between zero and unity, zero representing no similarity, and unity representing maximal similarity. It is calculated by the following equation:

$$SM = \frac{1}{1+l}, \quad (7)$$

where  $l$  is the Euclidian distance between the centroid of a given cluster (deterministic, chaotic, or random) and the point under consideration.

The second test involved filtering random model signals using a fourth order low-pass digital Butterworth filter. Our hypothesis was that as the random signals undergo low-pass filtering at gradually reduced cut-off frequencies, the level of randomness would be reduced, and the positioning of these new signals in relation to each of the deterministic, chaotic and random centroids, would change after being submitted to the DFNN algorithm. We expected that with the low-pass filtering of these signals with gradually decreasing cut-off frequencies, the level of randomness would drop, and the positioning of the signals would shift away from the random centroid towards the chaotic and deterministic centroids.

#### 2.6.4. Experiment with Electrogastrographic Signals to Detect Gastric Uncoupling

Gastric uncoupling is the loss of electrical synchronization in the stomach [18]. Since gastric motility is electrically controlled, such uncoupling may result in clinical complications such as gastroparesis [18]. In a canine experiment performed by Mintchev et al [18], gastric electrical uncoupling was artificially induced by surgically inhibiting the propagation of electric potentials throughout the length of the stomach using circumferential surgical cuts in the stomach physically separating sections of the organ. Electrogastrography (EGG) is a non-invasive method to record gastric electrical activity [19], and was utilized in this experiment in an attempt to validate its ability to recognize gastric electrical uncoupling. Three kinds of 8-channel EGG signals were recorded from 16 dogs: basal (B), after the first circumferential cut (FC), and after the second cut (SC), each representing three different levels of electrical desynchronization: (i) no uncoupling; (ii) mild induced uncoupling; and (iii) severe induced uncoupling. It has been shown that the amount of electrical uncoupling exhibited in the recorded signals increased with the number of circumferential cuts.

Utilizing the proposed DFNN algorithm, the B, FC, and SC signals from the EGG recordings were tested to assess how each of these signals positions itself with respect to the clusters of deterministic, chaotic, and random signals pre-identified in the experiments with the model signals. We hypothesized that since EGG signals were found to be chaotic [20], they would position themselves in the chaotic cluster of the plot, and that uncoupling will be detected by noticing that basal EGG signal patterns position closer to the deterministic cluster, while SC signal patterns position themselves closer to the random cluster.

Similarly to the model signals, each of the three types of EGG signals (B, FC, SC) was subjected to the DFNN algorithm. The resulting 128 wave-shape patterns for each state (8 EGG channels per state from each of the 16 dogs) were decomposed using wavelet analysis, and the mean and standard deviation were calculated for the coefficients of each EGG signal type with the aim to show how the B, FC, and SC signals position themselves with respect to the deterministic, chaotic and random regions defined using the model signals. The position of the B, FC, and SC signal patterns in each cluster were quantified using the same SM technique used to test the robustness of the DFNN algorithm [17].

---

### 3. Results

---

#### 3.1. DFNN Algorithm

The percent FNN up to a dimension of 32 were calculated for each model signal using a software package called Visual Recurrence Analysis (VRA) [21]. The calculated dimensions for each model signal resulted in unique wave-shapes (see examples in Figure 1). A total of 49 wave shapes were obtained: 9 from the deterministic model signals, 26 from the chaotic model signals, and 14 from the random model signals. Each of these wave-shapes was submitted to wavelet analysis decomposition using the Haar wavelet, the decomposition resulting in 32 coefficients per model signal. The mean and standard deviation of approximation coefficients per model signal were calculated, with a sample of the results shown in Table 1. A plot of the mean against the standard deviation for each model signal was built (Figure 3). Notice the tendency for the deterministic signals to cluster on the left

of the plot, the chaotic signals to cluster near the centre of the plot, and the random signals to cluster to the right of the plot.

To formalize the uniqueness of each group as representative of each model signal, the Fuzzy C-means algorithm was applied to the points of Figure 3. This resulted in a centroid being defined for each model signal group (graphically shown in Figure 3, numerically shown in Table 2), to clearly partition the deterministic, chaotic, and random model signals into specific regions of the plot.

### **3.2. Robustness of the Algorithm**

In the first test performed, the data points of two deterministic and chaotic signals were shuffled and the resulting signals submitted to the DFNN algorithm. All of the shuffled signals positioned themselves in the random region as expected (Figure 4).

The second test for robustness involved filtering random signals using a fourth order low-pass digital Butterworth filter. Three types of random signals were utilized (represented by exponential, uniform and normal probability density functions) and filtered at different normalized cut-off frequencies (0.8, 0.5, 0.3, 0.1, 0.01). With the filtering at decreasing cut-off frequencies, the signals shifted their position further away from the random centroid. This tendency is visualized in Figure 5.

### **3.3. Detection of Gastric Electrical Uncoupling**

Gastric electrical uncoupling as assessed by the DFNN algorithm can be demonstrated by a single representative point for each EGG signal type, calculated by obtaining the mean of the means and the standard deviations for each of the B, FC and SC signal wavelet coefficients (Figure 6a), resulting in three representative points shown in Figure 6b. Quantitatively, the calculations were performed using Equation 6, where the similarity measure SM was calculated for each of the B, FC, and SC signals with respect to the centroid of each of the deterministic, chaotic and random regions of the plot obtained from the model signals. The overall means and standard deviations for all SM calculations are shown in Table 3. Notice that the similarity of the B, FC and SC signals to the chaotic region was quite strong due to the high SM value, while their similarity to the deterministic and random regions was very weak due to a low SM value. Nevertheless, slight shift was noted towards the random centroid after the first circumferential cut (point FC on Figure 6b), and after the second cut the standard deviation of the obtained wavelet coefficients increased notably (point SC on Figure 6b).

Table 1 – Sample means and standard deviations of the wavelet coefficients obtained from some model signals.

Signal Type	Signal Model	Sample Mean	Sample Standard Deviation
<b>Deterministic</b>	Sine	3.2306	12.4698
	Triangular	2.6224	10.4899
	Square	0.0018	0.0071
	Sawtooth	0.0301	0.0071
<b>Chaotic</b>	Henon	36.7863	11.2378
	Ikeda	55.6975	2.2675
	Logistic	48.7745	19.1534
	Lorentz	40.1871	4.1175
	MGM	25.7374	11.1079
	Quadratic	37.8974	17.3651
<b>Random</b>	Impulse	70.3845	0.5949
	Normal	69.7791	0.6863
	Uniform	70.7354	0.6900

Table 2 - Centroid coordinates for each signal model group.

Signal Type	Centroid coordinate
<b>Deterministic centroid</b>	Mean: 3.3877 SD: 7.2483
<b>Chaotic centroid</b>	Mean: 40.9797 SD: 10.7933
<b>Random centroid</b>	Mean: 69.9643 SD: 1.8843

Table 3 – Means and standard deviations of the similarity measures from all EGG signals.

Signal Type	SM to Deterministic Cluster	SM to Chaotic Cluster	SM to Random Cluster
<b>Basal (B)</b>	Mean: 0.0304 SD: 0.0038	Mean: 0.1329 SD: 0.0449	Mean: 0.0274 SD: 0.0045
<b>First Cut (FC)</b>	Mean: 0.0293 SD: 0.0039	Mean: 0.1549 SD: 0.0586	Mean: 0.0278 SD: 0.0031
<b>Second Cut (SC)</b>	Mean: 0.0298 SD: 0.0039	Mean: 0.1546 SD: 0.0595	Mean: 0.0271 SD: 0.0025

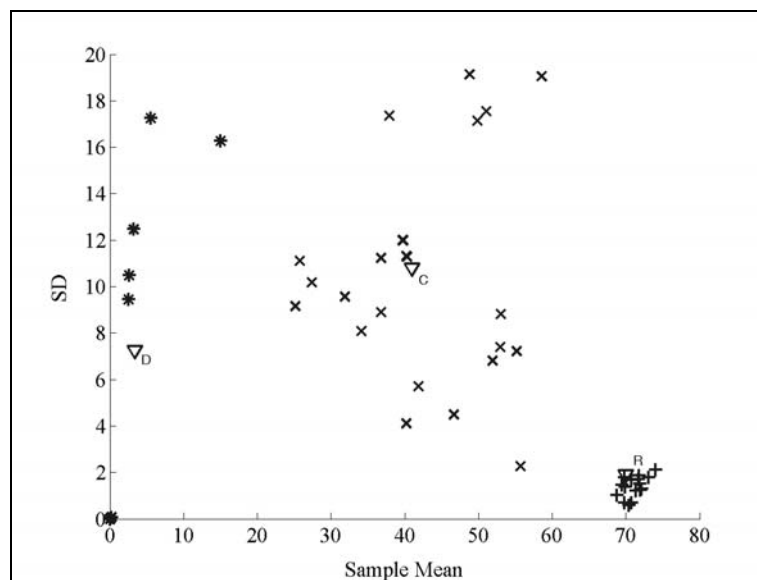


Figure 3 – Sample means and standard deviations of the wavelet coefficients obtained from deterministic (\*), chaotic (x), and random (+) model signals. The centroids ( $\nabla$ ) for each cluster were calculated using the Fuzzy C-Means algorithm.

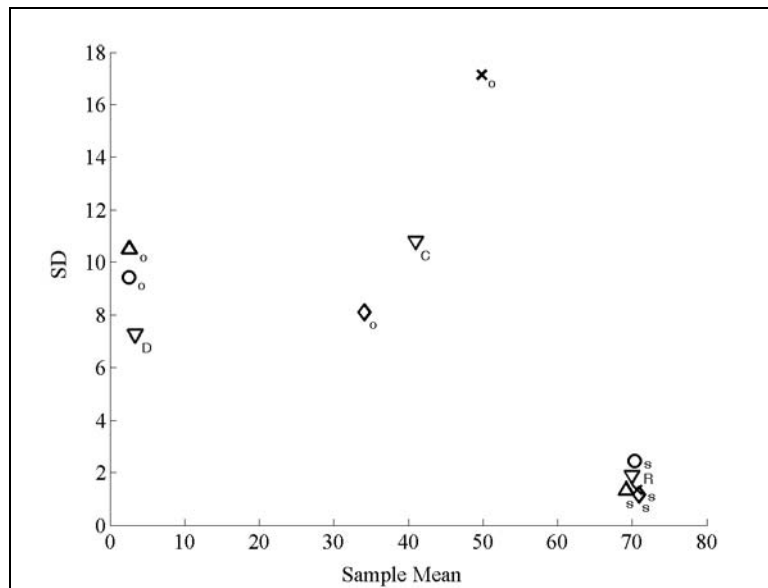


Figure 4 – Positions of the shuffled (s) and the original (o) deterministic (○ - sine wave, Δ - triangular wave) and chaotic (× - logistic map, ◇ - Lorenz map) signals.

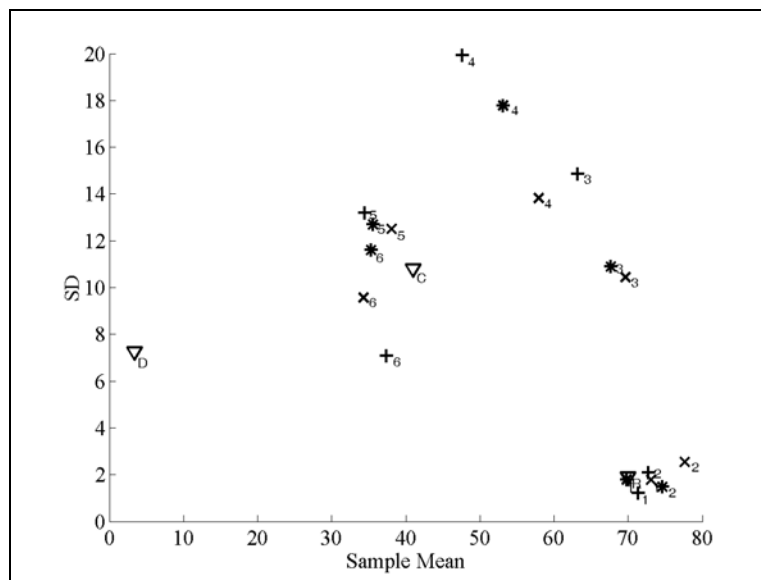


Figure 5 – Three random signals with different probability density functions, exponential (\*), uniform (+) and normal (×), filtered with a fourth order low-pass digital Butterworth filter at the following normalized cut-off frequencies: (1) no filtering, (2) 0.8, (3) 0.5, (4) 0.3, (5) 0.2, (6) 0.01.



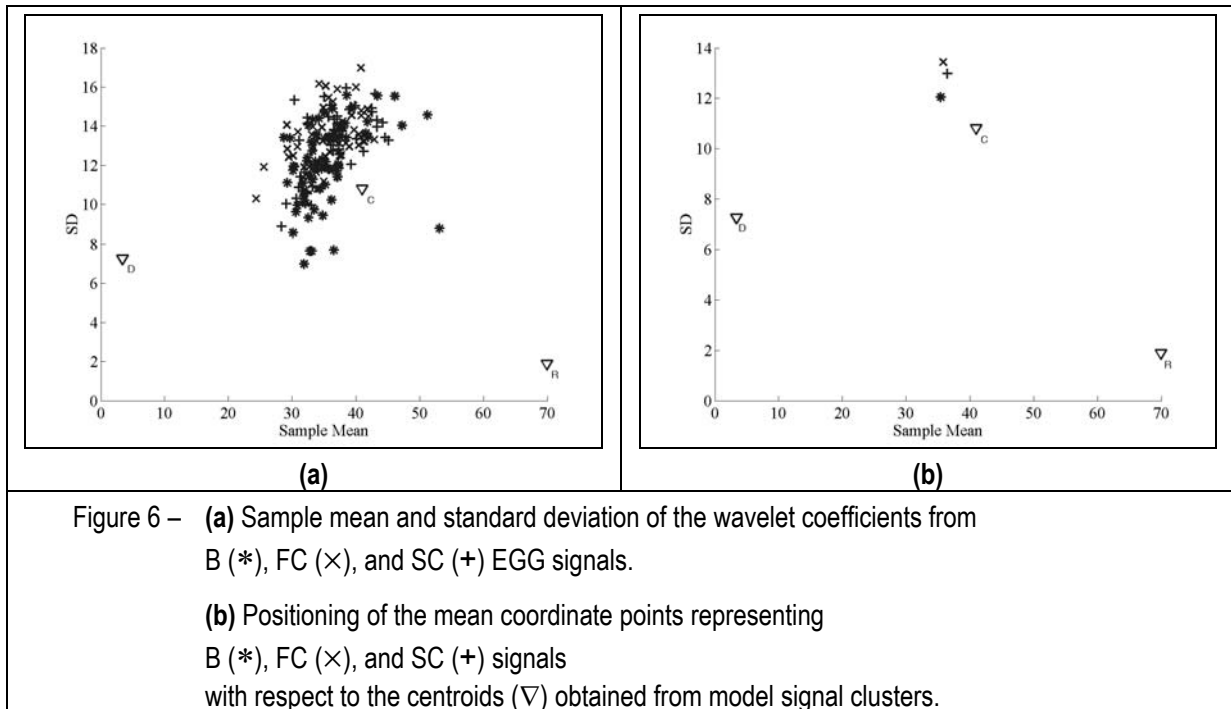


Figure 6 – (a) Sample mean and standard deviation of the wavelet coefficients from B (\*), FC (x), and SC (+) EGG signals.

(b) Positioning of the mean coordinate points representing B (\*), FC (x), and SC (+) signals with respect to the centroids ( $\nabla$ ) obtained from model signal clusters.

#### 4. Discussion

In the recent years chaos analysis of biomedical signals evolved into a powerful digital signal processing avenue [2, 4, 5, 22], often overshadowing the well established deterministic and stochastic signal processing tools [1]. However, a comprehensive methodology to determine the adequate digital signal processing tool set (deterministic, chaotic or stochastic) for a specific biomedical signal is still lacking.

In this study we developed an innovative procedure to examine whether given biomedical signals of interest belong to predefined clusters of deterministic, chaotic or random patterns obtained from model signals typical for each of these three categories. The intent was to algorithmically facilitate an informed quantitative decision on which signal processing tools were better suited for the processing of the biomedical signals under consideration. The proposed DFNN algorithm, combined with wavelet decomposition and subsequent statistical analysis were found to be excellent candidates for fulfilling this mission. The research was motivated by the observation that the shapes of the curves produced by the FNN algorithm appeared to be visually related to the signal type. Therefore, a pattern recognition technique based on a dyadic wavelet expansion of the FNN characteristic was developed. The method was tested on a selected set of artificially constructed signals, and then used to assess how some specific biomedical signals [23] position themselves in these categories. It is important to note that the method was tested on a selected set of model signals, and thus further testing with a variety of model signals might be appropriate to fully assess the capabilities and the limitations of the proposed technique.

The methodology resulted in a convenient and very clear clustering of deterministic, chaotic, and random signal patterns extracted from model signals (see Figure 3). Subsequent analysis of electrogastric signals in different states (basal, after mild invoked uncoupling, and after severe invoked uncoupling) confirmed previous suggestions that the EGG signals are inherently chaotic [20, 24]. Moreover, it was observed that the dominant chaotic nature of these signals, demonstrated by the fact that the DFNN algorithm resulted in their classification well in the middle of the predefined chaotic cluster (see Figure 4), most likely precluded a clear and significant shift from the basal pattern (B) when the signals recorded after the invoked uncouplings (FC and SC) were considered.

---

## 5. Conclusion

---

An innovative technique for classifying biomedical signals in three categories, deterministic, chaotic, and random was developed. The methodology was quantitatively tested using model signals belonging to each of these three categories, and actual electrogastrographic signals subjected to experimentally controlled uncoupling. The technique could be very useful in making an informed decision which digital signal processing toolset would be most appropriate for a specific type of biomedical signals.

---

## Acknowledgement

---

This study was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the Gastrointestinal Motility Laboratory (University of Alberta Hospitals) in Edmonton, Alberta, Canada

---

## References

---

- [1] R. E. Challis and R. I. Kitney, "Bio-Medical Signal-Processing (in 4 Parts). 1. Time-Domain Methods," *Medical & Biological Engineering & Computing*, vol. 28, pp. 509-524, 1990.
- [2] H. D. I. Abarbanel, *Analysis of observed chaotic data*. New York: Springer, 1996.
- [3] T. Gautama, D. P. Mandic, and M. M. Van Hulle, "A novel method for determining the nature of time series," *Biomedical Engineering, IEEE Transactions on*, vol. 51, pp. 728-736, 2004.
- [4] B. Chen and N. Wang, "Determining EMG embedding and fractal dimensions and its application," presented at Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE, 2000.
- [5] B. Huang and W. Kinsner, "Impact of low-rate sampling on the reconstruction of ECG in phase-space," presented at Electrical and Computer Engineering, 2000 Canadian Conference on, 2000.
- [6] F. Takens, "Detecting strange attractors in turbulence," in *Lecture notes in mathematics*, vol. 898, D. A. Rand and L. S. Young, Eds. Berlin: Springer, 1981, pp. 366-381.
- [7] W. Liebert and H. G. Schuster, "Proper choice of the time delay for the analysis of chaotic time series," *Physics Letters A*, vol. 142, pp. 107-111, 1989.
- [8] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, pp. 1134-1140, 1986.
- [9] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical reconstruction," *Physical Review A*, vol. 45, pp. 3403-3411, 1992.
- [10] A. M. Reza, "From Fourier Transform to Wavelet Transform, Basic Concepts," Spire Lab, UWM, White Paper October 27 1999.
- [11] M. Unser, "Wavelet Theory Demystified," *IEEE Transactions on Signal Processing*, vol. 51, pp. 470 - 483, 2003.
- [12] G. Chen, G. Chen, and R. J. P. de Figueiredo, "Feedback control of unknown chaotic dynamical systems based on time-series data," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on]*, vol. 46, pp. 640-644, 1999.
- [13] A. Ohsumi, H. Ijima, and T. Kuroishi, "Online detection of pulse sequence in random noise using a wavelet," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 47, pp. 2526-2531, 1999.
- [14] S. G. Mallat, *A Wavelet Tour of Signal Processing, 2nd ed.*: Academic Press, 1999.
- [15] P. P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on]*, vol. 48, pp. 1094-1109, 2001.
- [16] S. Miyamoto, "An overview and new methods in fuzzy clustering," presented at Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES '98. 1998 Second International Conference on, 1998.
- [17] D. S. Yeung and X. Z. Wang, "Using a neuro-fuzzy technique to improve the clustering based on similarity," presented at Systems, Man, and Cybernetics, 2000 IEEE International Conference on, 2000.
- [18] M. P. Mintchev, S. J. Otto, and K. L. Bowes, "Electrogastrography can recognize gastric electrical uncoupling in dogs," *Gastroenterology*, vol. 112, pp. 2006-2011, 1997.
- [19] M. A. M. T. Verhagen, L. J. Van Schelven, M. Samson, and A. J. P. M. Smout, "Pitfalls in the analysis of electrogastrographic recordings," *Gastroenterology*, vol. 117, pp. 453-460, 1999.

- 
- [20] J. Y. Carre, A. Høst-Madsen, K. L. Bowes, and M. P. Mintchev, "Analysis of the dynamics of the level of chaos associated with gastric electrical uncoupling in dogs," *Medical and Biological Engineering and Computing*, vol. 39, pp. 322-329, 2001.
- [21] E. Kononov, "Visual Recurrence Analysis (VRA)," 4.5 ed. <http://home.netcom.com/~eugenek>, 2003.
- [22] I. Yaylali, H. Kocak, and P. Jayakar, "Detection of seizures from small samples using nonlinear dynamic system theory," *Biomedical Engineering, IEEE Transactions on*, vol. 43, pp. 743-751, 1996.
- [23] C. Newton Price and M. P. Mintchev, "Quantitative evaluation of the dynamics of external factors influencing canine gastric electrical activity before and after uncoupling," *Journal of Medical Engineering and Technology*, vol. 26, pp. 239-246, 2002.
- [24] G. Lindberg, "Is the electrogastrogram a chaotic parameter?," In: Proceedings of the Fourth International Workshop on Electrogastrography, San Francisco, CA, May, 1996.
- 

### Authors' Information

---

**Charles Newton Price** – M.Sc. graduate student,  
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4

**Renato J. de Sobral Cintra** – Ph.D. graduate student,  
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4  
Department of Electronics & Systems, Federal University of Pernambuco; Recife, Pernambuco, Brazil

**David T. Westwick** – Associate Prof., Dr.,  
Department of Electrical and Computer Engineering; University of Calgary, Calgary, Alberta, Canada T2N 1N4

**Martin P. Mintchev** – Prof., Dr.,  
Department of Electrical and Computer Engineering; University of Calgary; 2500 University Drive NW; Calgary, Alberta, Canada T2N 1N4  
Department of Surgery, University of Alberta; Edmonton, Alberta, Canada T6G 2B7  
phone: (403) 220-5309; fax: (403) 282-6855; e-mail: [mintchev@enel.ucalgary.ca](mailto:mintchev@enel.ucalgary.ca)

## TRACKING SENSORS BOTTLENECK PROBLEM SOLUTION USING BIOLOGICAL MODELS OF ATTENTION

**Alexander Fish, Orly Yadid-Pecht**

**Abstract:** *Every high resolution imaging system suffers from the bottleneck problem. This problem relates to the huge amount of data transmission from the sensor array to a digital signal processing (DSP) and to bottleneck in performance, caused by the requirement to process a large amount of information in parallel. The same problem exists in biological vision systems, where the information, sensed by many millions of receptors should be transmitted and processed in real time. Models, describing the bottleneck problem solutions in biological systems fall in the field of visual attention. This paper presents the bottleneck problem existing in imagers used for real time salient target tracking and proposes a simple solution by employing models of attention, found in biological systems. The bottleneck problem in imaging systems is presented, the existing models of visual attention are discussed and the architecture of the proposed imager is shown.*

**Keywords:** *Bottleneck problem, image processing, tracking imager, models of attention*

---

## 1. Introduction

---

Driven by the demands of commercial, consumer, space and security applications, image sensors became a very hot topic and a major category of high-volume semiconductor production [1]. This is due to the imminent introduction of imaging devices in high volume consumer applications such as cell phones, automobiles, PC-based video applications, "smart" toys and, of course, digital still and video cameras. While most of consumer applications are satisfied with relatively low-resolution imagers, image sensors used for target tracking in space, navigation and security applications require high spatial resolution. In addition, these tracking sensors are usually supposed to provide real time tracking after multiple targets in the field of view (FOV), such as stars, missiles and others. The demands for high resolution and real time performance result in a bottleneck problem relating to the large amount of information transmission from the imager to the digital signal processing (DSP) or processor and in bottleneck in performance, caused by the requirement to process a large amount of information in parallel. The simple solution to the bottleneck in performance is to use more advanced processors to implement the required tracking algorithms or to use dedicated hardware built specially for the required algorithm implementation [2]. However, the solution to the performance bottleneck still does not relax the requirement for the large data transmission between sensor and processor.

The same problem exists in biological vision systems. Compared to the state-of-the-art artificial imaging systems, having about twenty millions sensors, the human eye has more than one hundred million receptors (rods and cones). Thus, the question is how the biological vision systems succeed to transmit and to process such a large amount of information in real time? The answer is that to cope with potential overload, the brain is equipped with a variety of attentional mechanisms [3]. These mechanisms have two important functions: (a) attention can be used to select relevant information and/or to ignore the irrelevant or interfering information; (b) attention can modulate or enhance the selected information according to the state and goals of the perceiver. Numerous research efforts in physiology were triggered during the last five decades to understand the attention mechanism [4-12]. Generally, works related to physiological analysis of the human attention system can be divided into two main groups: those that present a spatial (spotlight) model for visual attention [4-6] and those following object-based attention [7-12]. The main difference between these models is that the object-based theory is based on the assumption that attention is referenced to a target or perceptual groups in the visual field, while the spotlight theory indicates that attention selects a place at which to enhance the efficiency of information processing.

The design of efficient real time tracking systems mostly depends on deep understanding of the model of visual attention [12-14]. This paper briefly describes spotlight and object-based models of attention and proposes a solution for the bottleneck problem in image systems for salient targets tracking based on the study and utilization of the spatial (spotlight) model of attention. Two possible sensor architectures are presented and discussed.

Section 2 briefly describes the bottleneck problem in high resolution imaging systems. A review of existing models of attention is presented in Section 3. Section 4 presents descriptions of two sensor architectures, comparing it with the existing spatial (spotlight) models of attention. Section 5 concludes the paper.

---

## 2. The Bottleneck Problem in High Resolution Image Systems

---

As mentioned in section 1, two bottlenecks exist in high-resolution image systems: the data transmission bottleneck and performance bottleneck. The solution for the bottleneck in performance relates to increasing the processing power. This can be performed in the following ways: (a) to use more advanced processors to implement the algorithms, (b) to use dedicated hardware built especially for the required algorithms implementation and (c) employing more than one DSP/processor. Although these solutions are expensive and can dramatically increase the cost of the whole system, they provide simple and trustworthy solutions. Fig. 1 shows an example of such a kind of an imaging system. As can be seen, the system consists of the image sensors array (can either be implemented as a Charge Coupled Device (CCD) or as a standard CMOS imager, as will be described below), a number of processors (or DSPs) and a memory.

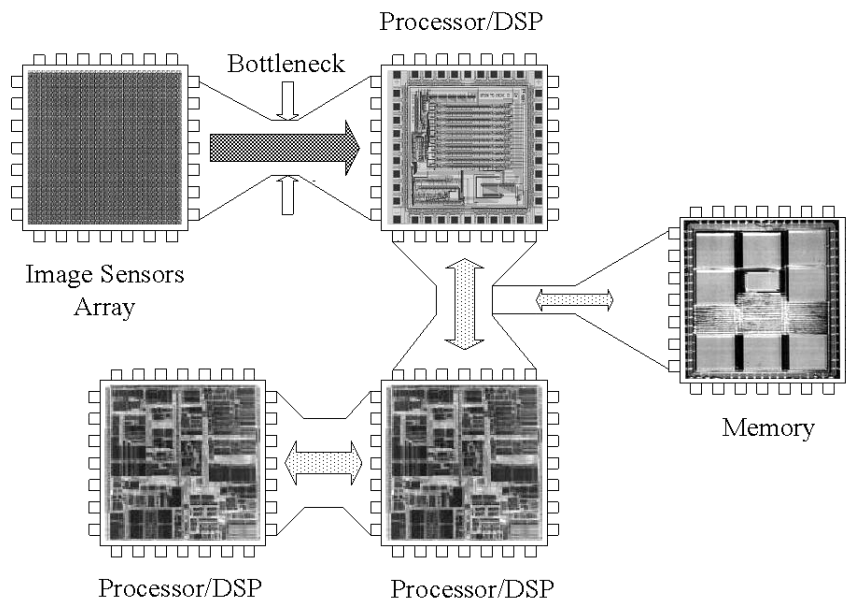


Fig.1. An example of a typical image system, incorporating an image sensor array, processors/DSPs for image processing and memory

The data transmission bottleneck from the image sensors array to the processor/DSP (see Fig. 1) is more difficult for solution. The most efficient way to solve the problem is on-chip implementation of some algorithms that can relax the information bottleneck by reducing the amount of data for transmission. However, this solution is almost impossible in CCD sensors which cannot easily be integrated with standard CMOS analog and digital circuits due to additional fabrication complexity and increased cost. On the other hand, CMOS technology provides the possibility for integrating imaging and image processing algorithms functions onto a single chip, creating so called "smart" image sensors. Unlike CCD image sensors, CMOS imagers use digital memory style readout, using row decoders and column amplifiers. This readout allows random access to pixels so that selective readout of windows of interest is allowed. In this paper all further discussions will be related to CMOS image sensors.

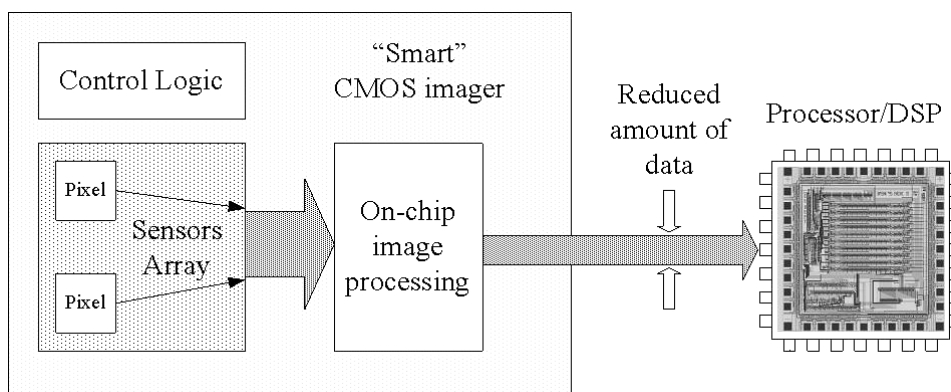


Fig.2. An example of an imaging system, employing a "smart" CMOS image sensor with on-chip processing and processors/DSPs for image processing

Fig. 2 shows an example of an imaging system employing a "smart" CMOS image sensor and a single processor/DSP. As can be seen, in this system the image processing can be performed at three different levels: (a) at the pixel level – CMOS technology allows insertion of additional circuitry into the pixel, (b) on-chip image processing and, of course (c) image processing by processor/DSP. The larger amount of processing performed in the first two levels, the less amount of information necessary to be transmitted from the CMOS imager to the DSP. In addition, smaller computation resources are required.

Generally, image processing at early stages (pixel level and on-chip level) can solve both data transmission bottleneck and performance bottleneck problems. Tracking imager architectures, proposed in this paper, will try to imitate the models of attention emphasizing on the requirement to perform the most of processing at early stages.

### 3. Visual Models of Attention

Although many research efforts were triggered during the last decades and numerous models of attention have been proposed over the years, there is still much confusion as to the nature and role of attention. Generally two models of attention exist: spatial (spotlight) or early attention and object-based, or late attention. While the object-based theory suggests that the visual world is parsed into objects or perceptual groups, the spatial (spotlight) model purports that attention is directed to unparsed regions of space. Experimental research provides some degree of support to both models of attention. While both models are useful in understanding the processing of visual information, the spotlight model suffers from more drawbacks than the object-based model. However, the spotlight model is simpler and can be more useful for tracking imager implementations, as will be shown below.

#### 3.1 The Spatial (Spotlight) Model

The model of spotlight visual attention mainly grew out of the application of information theory developed by Shannon. In electronic systems, similar to physiological, the amount of the incoming information is limited by the system resources. There are two main models of spotlight attention. The simplest model can be looked upon as a spatial filter, where what falls outside the attentional spotlight is assumed not to be processed. In the second model, the spotlight serves to concentrate attentional resources to a particular region in space, thus enhancing processing at that location and almost eliminating processing of the unattended regions. The main difference between these models is that in the first one the spotlight only passively blocks the irrelevant information, while in the second model it actively directs the "processing efforts" to the chosen region.

Fig. 3(a) and Fig 3(b) visually clarify the difference between the spatial filtering and spotlight attention.



Fig.3 (a). An example of spatial filtering

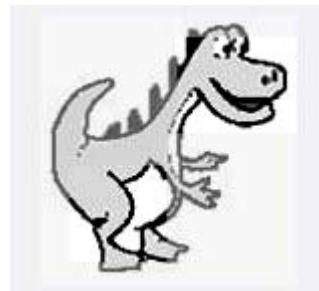


Fig.3 (b). An example of spotlight model of attention

A conventional view of the spotlight model assumes that only a single region of interest is processed at a certain time point and supposes smooth movement to other regions of interest. Later versions of the spotlight model assume that the attentional spotlight can be divided between several regions in space. In addition, the latter support the theory that the spotlight moves discretely from one region to the other.

#### 3.2 Object-based Model

As reviewed above, the spotlight metaphor is useful for understanding how attention is deployed across space. However, this metaphor has serious limitations. A detailed analysis of spotlight model drawbacks can be found in [3]. Object-based attention model suit to more practical experiments in physiology and is based on the assumption that attention is referred to discrete objects in the visual field. However being more practical, in contrast to the spotlight model, where one would predict that two nearby or overlapping objects are attended as a single object, in the object-based model this divided attention between objects results in less efficient processing than attending to a single object. It should be noted that spotlight and object-based attention theories are not contradictory but rather complementary. Nevertheless, in many cases the object-based theory explains many phenomena better than the spotlight model does.

The object-based model is more complicated for implementation, since it requires objects' recognition, while the spotlight model only requires identifying the regions of interest, where the attentional resources will be concentrated for further processing.

#### 4. The Proposed Architecture for the CMOS Tracking Imager

In this Section two possible architectures of tracking CMOS imagers are discussed. While these architectures seem similar, the first one employing the spatial filtering model and the second one employing the spotlight attention model. The operation of both imagers will be described with the reference to an example of input scene in FOV, as shown in Fig. 4.



Fig.4 An example of input scene in FOV

In this scene three different types of regions can be observed: (a) two regions consisting of large salient targets (stars), (b) a number of regions consisting of small salient stars and (c) regions that don't consist of any targets of interest. The observer is usually interested in tracking the targets mentioned in group (a), but sometimes there is interest in targets both from groups (a) and (b). Moreover, sometimes the observer is interested in tracking the targets from group (b) and the targets from group (a) serve as salient distractors. Note, the term real time tracking relates to the ability to calculate the center of mass (COM) coordinates of the tracked target in real time.

##### 4.1 The Spatial Filtering Based Architecture

Fig. 5 shows the architecture of the CMOS tracking image system based on the spatial filtering model.

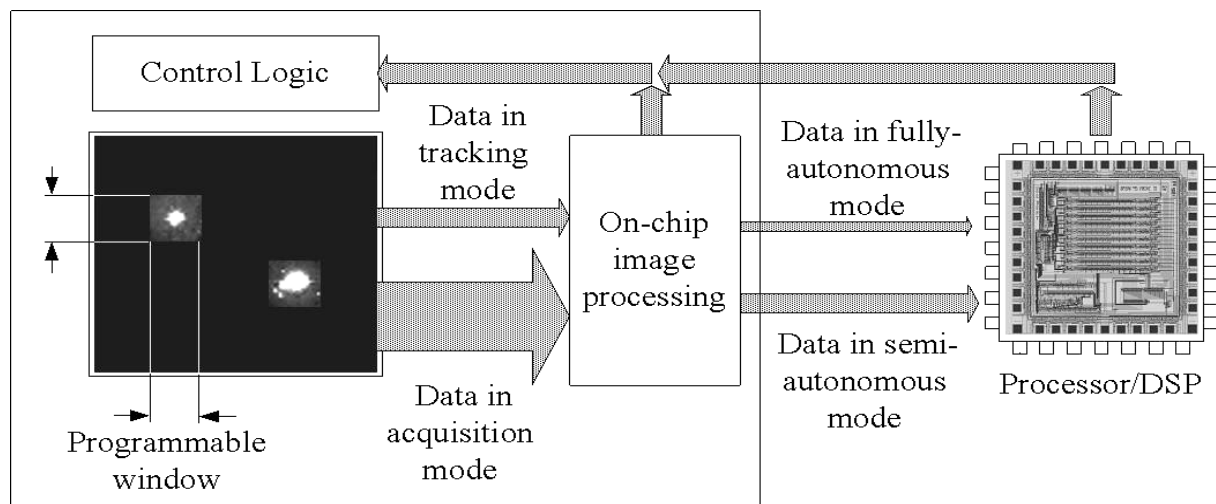


Fig.5 Architecture of the CMOS tracking image system based on the spatial filtering model

The proposed sensor has two modes of operation: the fully autonomous mode and the semi-autonomous. In the fully autonomous mode all functions required for target tracking are performed by the sensor (at the pixel level and by on-chip image processing). The only data transmitted to the processor/DSP in this case is the tracked targets coordinates. This mode is very efficient by means of bottleneck problems solution; however it allows less flexibility and influence on the tracking process by the user. In the semi-autonomous mode, part of the functionality is performed in the circuit level and the chip level and part of processing is done by the processor/DSP. In this case feedback from the processor/DSP to the sensor exists and more flexibility

is achieved; however more data flows from the sensors to the processor/DSP and back and more processor/DSP resources are used to complete the real time tracking (see Fig. 5).

The real time targets tracking is accomplished in two stages of operation: target acquisition and target tracking. In the acquisition mode  $N$  most salient targets of interest (the number of targets  $N$  can be predefined by the systems or can be user-defined) in the FOV are found. Then,  $N$  windows of interest with programmable size around the targets are defined, using the control logic block. These windows define the active regions, where the subsequent processing will occur, similar to the flexible spotlight size in the biological systems. In the tracking stage, the system sequentially attends only to the previously chosen regions, while completely inhibiting the dataflow from the other regions. This way the system based on the spatial model of attention allows distractors elimination, oppositely to a case of the spotlight model. According to the spotlight model appearance of the additional "salient" targets during the tracking of given targets of interest causes temporary or even permanent loss of the desired target.

Thanks to the control logic and to the CMOS imager flexibility, the proposed concept permits choosing the attended regions in the desired order, independent on the targets saliency. In addition it allows shifting the attention from one active region to the other, independent of the distance between the targets.

As can be seen more information is transmitted from the sensor array to the on-chip system-processing block during the acquisition mode than during the tracking mode. The reason for this is that during the acquisition mode the whole image is captured and transmitted to the on-chip image-processing block for further processing. In case of the fully autonomous mode of operation, the on-chip processing finds the center of mass coordinates of all targets of interest and windows of interest are defined by the control logic. In a case of semi-autonomous mode, some processing is performed in the on-chip processing block and the data is transmitted to the processor/DSP for further processing and windows of interest definition. Note, the acquisition mode is required only once at the beginning of tracking. During the tracking mode, only information from the chosen windows of interest is transmitted, dramatically reducing the bottleneck problem between the sensor and on-chip image-processing block and between the imager chip to processor/DSP.

#### 4.2 The Spotlight Based Architecture

Fig. 6 shows the architecture of the CMOS tracking image system based on the spotlight model. The principle of this system is very similar to the concept, presented in sub-section 4.1.

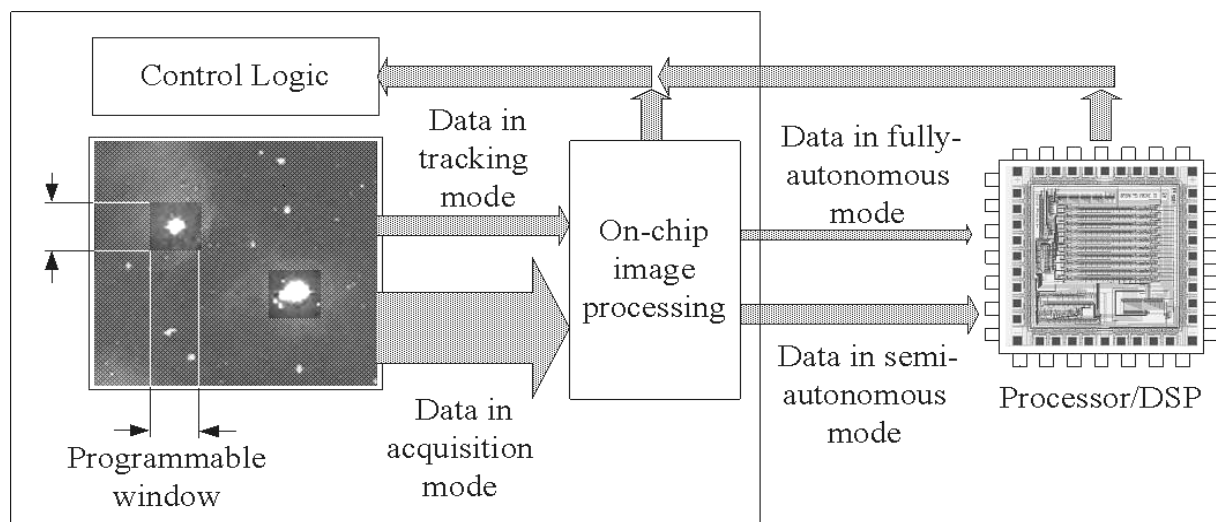


Fig.6 Architecture of the CMOS tracking image system based on the spotlight model

However, there is one important difference. While the spatial filter based imager filters all information that does not fall into the windows of interest, the spotlight based system, only reduces the amount of information transmitted from these regions for further processing. This is performed by a unique feature of CMOS image sensors that allow implementation of adaptive multiple resolution sensors [15]. As can be seen in Fig. 6, the two



most important regions of interest are captured with full resolution (in the same way like in the system presented on sub-section 4.1), while all other regions are captured with reduced resolution. On one hand this allows to change the defined windows of interest according to the events in the FOV. On the other hand, this architecture is still significantly reduces the amount of information transmission from the sensor.

---

## 5. Conclusions

---

Two architectures of tracking image systems were proposed. Both the data transmission bottleneck and the performance bottleneck are reduced in the proposed imagers due to employing the spatial filtering and spotlight models of attention, found in biological systems. The proposed imagers can be easily implemented in a standard CMOS technology. Both imagers can operate in the full autonomous and semi-autonomous modes of operation. A brief description of the spatial and object-based models of attention was presented and an explanation of the proposed image systems operation was provided. Further research includes implementation of the proposed sensors in an advanced CMOS technology.

---

## Acknowledgements

---

We would like to thank Alexander Spivakovsky and Evgeny Artyomov for their helpful suggestions during the preparation of this work.

---

## Bibliography

---

- [1] O. Yadid-Pecht, R. Etienne-Cummings. CMOS imagers: from phototransduction to image processing. Kluwer Academic Publishers, 2004.
- [2] R. William, H. Engineering. Using FPGAs for DSP Image Processing. FPGA journal, available [http://www.fpgajournal.com/articles/imaging\\_hunt.htm](http://www.fpgajournal.com/articles/imaging_hunt.htm)
- [3] Chun, M. M., & Wolfe, J. M. Visual Attention. In E. B. Goldstein (Ed.), Blackwell's Handbook of Perception, Vol. Ch 9, pp. 272-310. Oxford, UK: Blackwell, 2001.
- [4] R. W. Remington, L. Pierce. Moving attention: Evidence for Time-invariant shifts of visual selection attention. Perception and Psychophysics, vol. 35, pp. 393-399, 1984.
- [5] J. Moran, R. Desimone. Selective attention gates visual processing in the Extrastriate Cortex. Science, vol. 229, pp. 784-787, 1985.
- [6] G. Sperling, E. Weichselgartner. Episodic theory of the dynamics of spatial attention. Psychological review, vol. 102, no. 3, pp. 503-532, 1995.
- [7] A. Treisman. Features and targets: The fourteenth Barlett memorial lecture. Quarterly Journal of Experimental Psychology, 40A, pp. 201-237, 1988.
- [8] A. Treisman. Feature binding, attention and target perception. Philosophical Transactions of the Royal Society of London, vol. 353, pp. 1295-1306, 1998.
- [9] A. Treisman, G. Gelade. A feature-integration theory of attention. Cognitive Psychology, vol. 12, pp. 97-136, 1980.
- [10] A. Treisman, D. Kahneman, J. Burkell. Perceptual targets and the cost of filtering. Perception & Psychophysics, vol. 33, pp. 527-532, 1983.
- [11] S. Yantis. Targets, attention, and perceptual experience in "Visual Attention". R. D. Wright (Eds.), pp. 187-214. Oxford, NY: Oxford University Press, 1998.
- [12] M. I. Posner. Orienting of attention. Quarterly Journal of Experimental Psychology, vol. 32, pp. 3-25, 1980.
- [13] C. Koch, B. Mathur. Neuromorphic vision chips. IEEE Spectrum, vol. 33, pp. 38-46, May 1996
- [14] C. Koch, S. Ullman. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. Human Neurobiology, vol. 4, pp. 219-227, 1985.
- [15] E. Artyomov, O. Yadid-Pecht. Adaptive Multiple Resolution CMOS Active Pixel Sensor. ISCAS 2004, Vancouver, Canada, May 2004, Vol. IV, pp. 836-839.

---

## Authors' Information

---

**Alexander Fish** – The VLSI Systems Center, Ben-Gurion University, Beer Sheva, Israel; e-mail: [afish@ee.bgu.ac.il](mailto:afish@ee.bgu.ac.il)

**Orly Yadid-Pecht** – The VLSI Systems Center, Ben-Gurion University, Beer Sheva, Israel or Dept of Electrical and Computer Engineering, University of Calgary, Alberta, Canada; e-mail: [oyp@ee.bgu.ac.il](mailto:oyp@ee.bgu.ac.il)

## NEURAL CONTROL MODEL OF CHAOTIC DYNAMIC SYSTEMS

**Cristina Hernández de la Sota, Juan Castellanos Peñuela,  
Rafael Gonzalo Molina, Valentín Palencia Alejandro**

**Abstract:** *The last decades have seen a dramatic growth, and many important theoretical advances, in the field of dynamic systems control. Artificial neural networks have been used to control nonlinear systems since they are able to compute complex functions concerning nonlinear decision. This work presents neural network architecture based on the backpropagation that can be used as controllers in order to stabilize unsteady periodic orbits. It also presents a neural network based method for transferring the dynamics among attractors, giving a more efficient system control. The procedure may be applied to every point of the basin, no matter how far away from the attractor they are. Finally, this work shows how two mixed chaotic signals can be controlled using a backpropagation neural net as filter, in order to separate and control both of them at the same time. The neural network provides a more effective control; it can be applied to the system at any point, even being too far from the desired state, avoiding long transient times. The control can be applied if there are only a few data of the system, and it will remain stable much more time even with small random dynamic noise. The net achieves a more effective control, improving the troubles arising with classical feedback methods. Moreover, the system computes a solution starting from any point, even being far away from the desired one, avoiding delays. Also with a few amount of data, the connectionist system can be applied, remaining stable during a long time even with small random dynamic noise.*

**Keywords:** *neural network, backpropagation, dynamic systems control, feedback methods.*

---

### Introduction

---

During last decades, the physicists, astronomers and economists created a way to understand the development of complexity in nature. The new science named chaos offers a method to see order and guideline where before you could only observe the chance, the irregularity and the chaos. The chaos goes beyond the traditional scientific disciplines, being the science of the global nature of the systems, it has got together thinkers from very distant fields, from weather turbulences to the complicated beats of human heart, and from the design of the snowflakes to the sand whirlwinds of the desert.

The chaotic phenomena takes place everywhere as much in natural systems as in mechanisms constructed by man. Recent works have been centered mainly in describing and characterizing the chaotic behavior in situations where there is no human intervention. The control of chaotic signals is one of the most relevant search areas that have appeared during the last years. Recently, ideas and techniques have been proposed to turn chaotic orbits into desired periodic orbits, using controls temporarily programmed. For example, these techniques have been applied in mechanical systems [SINH98], laser [ROYR92], circuits [JOHN95], chemical reactions [PETR94], biological systems [GARF92] [BRAN95], etc.

However, the used methods of control have several disadvantages for their application, it is necessary to have enough data, the control is applied only when the state of the system is very close to the desired state, producing great transitory times closely together before activating the control [BAR95], the control is only effective in points next to the desired state and after a time the controlled orbit is destabilized, due to the computational error accumulated.

In this article, neural networks are designed to be used as controller for chaotic dynamic systems, overcoming the problems that appear when using another type of controllers.

---

### Model of Neural Control

---

The growing interest in neural networks composed of simple processing elements (neurons) has led to a wide use of such networks to control learning of dynamic systems [SONT93]. The capacity of the networks to produce a generalization and an efficient adaptation makes them excellent candidates for the control of linear as much

as non-linear dynamic systems. The objective of a controller based on neural networks is generating a correct control of the signal, to direct the dynamic from the initial state to the desired final state. The located execution and the facility to make a controller with a network depend mainly on the algorithm of the chosen learning, as well as on the architecture used for the control. In most of the designs, the backpropagation is used as learning algorithm.

The objective of this work is the use of neural networks as a structure of generic model for the identification and control of chaotic dynamic systems. The procedure that must be executed to control a chaotic dynamic system is shown in figure 1.

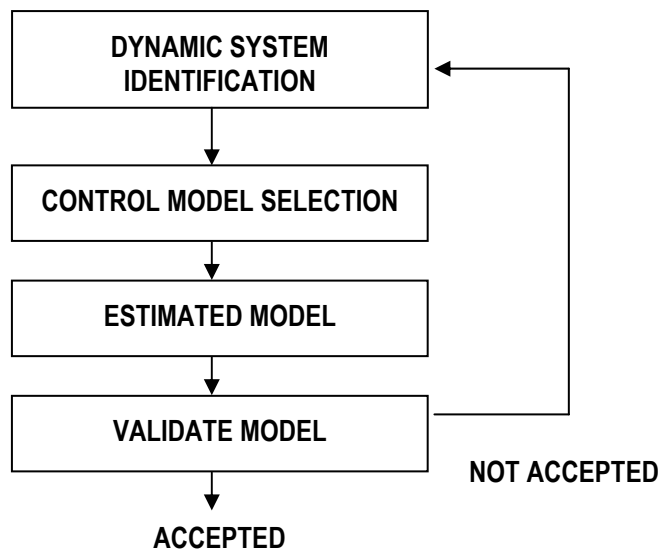


Figure 1: Design procedure of a control model

- Identification of the Chaotic Dynamic System

This phase consists in identifying the system, describing the fundamental data, the operative region, and the selection of the pattern.

$$Z_n = \{[u(t), y(t)]/t = 1 \dots N\}$$

where  $\{u(t)\}$  is the set of inputs, in other words, the signal that is desired to control,  $\{y(t)\}$  represents the output signal,  $t$  represents the moment of the pattern. If the considered system has more than an input/output,  $u(t)$ , and  $y(t)$  are vectors.

- Selection of the Control Model

Once the data set is obtained, the next step is to select a structure for the control model. It is necessary to choose a set of input patterns, but it is also required the architecture of the neural network. After defining the structure, the next step is to decide the input patterns and the number of them used to train the network.

- Estimated Model

Now it will be investigated what steps are necessary to make the control effective and to guarantee the convergence of the trajectories towards the desired orbit. The control is effective if there is some  $\delta > 0$  and  $t_0$  so that for  $t > t_0$  the distance between the trajectory and the stabilized periodic orbit is minor than  $\delta$ .

- Validated Model

When training a network, the following step is to evaluate it, to study the final errors. The most common validation method is to investigate residual (error prediction) by means of crossed validations of a set of tests. The visual inspection of the prediction graph compared with the wished output is probably the most important tool.

## Identification of the Chaotic Dynamic System

The systems that are going to be controlled are the non-linear and chaotic dynamic systems that depend on a system of parameters,  $p$ . The basic function is:  $\frac{dx(t)}{dt} = F(x(t), p)$ ; where  $F: \mathfrak{R}^n \rightarrow \mathfrak{R}^n$  is a continuous function.

The other type of systems is the discrete dynamic system, represented by an equation in non-linear differences. They are described like a function  $f: X \rightarrow X$  that determines the behavior or evolution of the set when time moves forward. The control system inputs are the orbits of the elements, the orbit of  $x \in X$  is defined like the succession  $x_0, x_1, x_2, \dots, x_n, \dots$  achieved by means of the rule:  $x_{n+1} = f(x_n)$  with  $x_0 = x$

The points of the orbit are obtained:

$$x_1 = f(x_0) = f(x); x_2 = f(x_1) = f(f(x)) = f^2(x); x_3 = f(x_2) = f(f^2(x)) = f^3(x); \dots x_n = f(f(\dots f(x)\dots)) = f^n(x) \text{ n times}$$

The behavior of the orbits can be very varied, depending on the dynamics of the system.

The objective is to control the dynamic system in some unstable periodic orbit or limit cycle that is within the chaotic attractor. Therefore, the output of the system will be the limit cycle of period-1 or greater in which the system must be controlled. In order to find the outputs it is necessary to consider:

- A point  $\alpha$  is an attractor for the function  $f(x)$  if there is a neighborhood around  $\alpha$  so that the point orbits in the neighborhood converge to  $\alpha$ . In other words, if values are taken near to  $\alpha$  the orbits will converge to  $\alpha$ .
- The simplest attractor is the fixed point. A point  $\alpha$  is a fixed point for the function  $f(x)$  if  $f(\alpha) = \alpha$
- A point  $\alpha$  is periodic if a positive integer number  $\tau$  exists so that  $f^\tau(\alpha) = \alpha$  and  $f^i(\alpha) \neq \alpha$  for  $0 < i < \tau$ . The integer  $\tau$  is known as the period of  $\alpha$
- Being  $\alpha$  a fixed point attractor; the set of initial values  $x_0$  whose orbits converge to  $\alpha$ , form the basin of attraction of  $\alpha$

The discrete systems that have been controlled are:

### Discrete Systems:

Systems  $f: R^2 \rightarrow R^2$  are controlled, non-linear discrete systems of second order, all the trajectories are focused towards the stable point  $x_{n+1} = f(x_n)$ , where  $x_n \in R^2$ . The chosen systems are Henon [MART95], Lozi [CHEN92], Ikeda [CASD89], and Tinkerbell [NUSS97]. The same type of previous discrete systems is controlled in unstable periodic orbits (limit cycle); the systems are Ikeda and Tinkerbell.

Another approach to control the chaos is referred to as the so-called entrainment and migration control methods, proposed by Jackson [JACK91] and Hübler [HUBL89]. To accomplish the migration control, the discrete system of Gumowski and Mira is used. This system has several attractors and several bounded convergent regions in the basin of attraction.

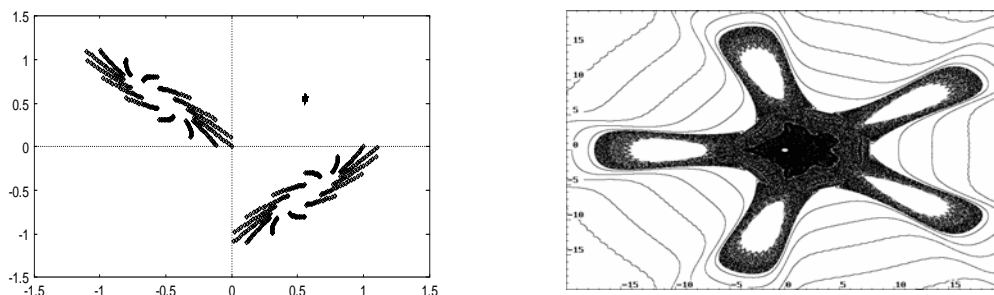
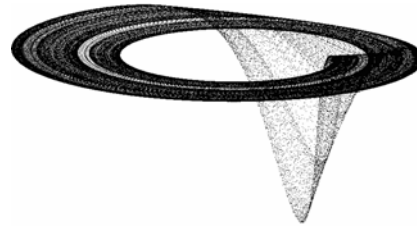


Figure. 2. Lozi system trajectory and Gumowski/ Mira map

**Continuous Systems:** Continuous systems  $f : R^3 \rightarrow R^3$  are controlled in an orbit of period 1, that is to say, in an equilibrium point. The system is the one of Lorenz [GULI92], and the Rössler system.



Figure 3. Lorenz's Attractor



Rössler's Attractor

---

### Selection of the Control Model

---

The structure of the Model has been divided into two sub problems:

- To design the network architecture
- To choose the structure of the input patterns

#### Architecture of the Neural Network

According to Lippmann [LIPP87], to characterize a model of neural network it must specify:

- The Transference Function of each node
- The Topology of the Network. This is defined by the number of nodes and the set of interconnections between them.
- The Rules of learning. They are the rules that regulate the form to find the weights associated to the connections.

**The Transference Function.** The activation function of the neurons is the sigmoid.

**The Topology of the neural network** The Neural Network employed as main controller consisting of three layers of neurons (input layer, hidden layer and output layer).

*Input layer:*

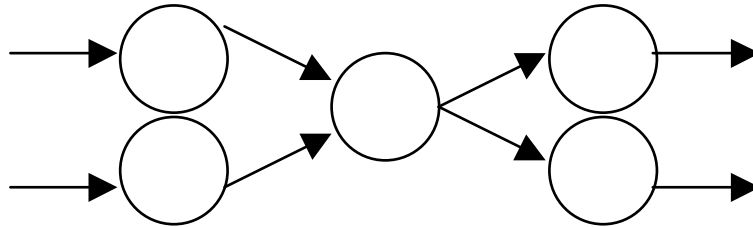
- When the control is made in a equilibrium point; the input layer has two neurons, one for each of the variable of the function  $f$  that is going to be controlled for the discrete functions  $f : R^2 \rightarrow R^2$  and three neurons, one for each of the variable of the function  $f$  that is going to be controlled for the continuous functions  $f : R^3 \rightarrow R^3$
- When the control is made in a limit cycle, the components of the vector are separated from the function  $f$ , first it will be applied to the first component of the function and next to the second. The input layer will have as many neurons as the period of the limit cycle. If the control is made in a limit cycle of period 7, the input layer has 7 neurons; if the period is of 5, the input layer has 5 neurons.

*Output layer:*

- When the control is made in an equilibrium point; the output layer has two or three neurons corresponding to the coordinates of the stable point.
- When the control is made in a limit cycle, the output layer will have as many neurons as the period of the limit cycle; if the period of the output layer is of seven neurons it will have 7 neurons that correspond to the first coordinates of period-7 point.

The number of neurons in the hidden layer plays an important role in the learning performance and generalization capability of the network:

- For the discrete functions  $f : R^2 \rightarrow R^2$  the hidden layer will have one hidden neuron when the control is made in an equilibrium point.



**Figure 4.** Neural Network architecture to control chaotic signals (with only one hidden neuron)

- For the continuous functions  $f : R^3 \rightarrow R^3$ , several simulations have been performed in order to know how the number of hidden neurons affects the mean square error in finding the stable point.

**The Rules of Learning.** The algorithm that is going to be used to adjust the weights is backpropagation, which follows a descent according to the gradient that minimizes the error function.

### Structure of the input patterns

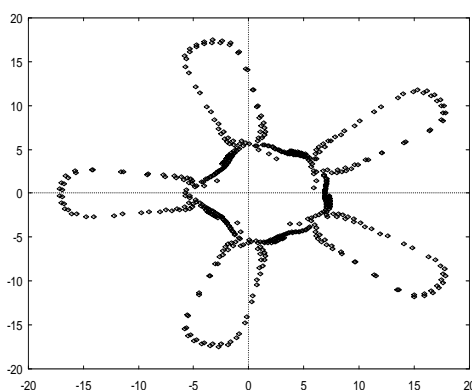
The input patterns are necessary to define appropriately it, to avoid falling in one of the greater problems: a local minimum. The patterns to train the network will be formed by system orbits, obtained starting from a point that is within the basin of attraction of the limit cycle chosen for controlling the system.

#### **1. Input Patterns.**

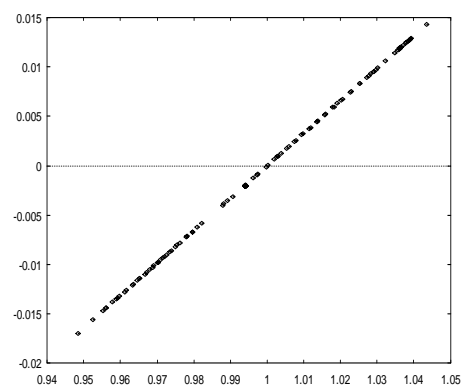
The input patterns are obtained taking a starting point  $(x_0, y_0)$  and finding the temporary series from the components of the function, iterating  $x_{n+1} = f(x_n)$ . The input file is constructed from a point, finding the temporary series with 500 patterns.

When the control is made in a limit cycle, for example, a period-5 limit cycle, input patterns will be constructed in the following way: From a point  $A = (x_0, y_0)$  the first pattern will be constructed:

$\{x_0, f(x_0), f^2(x_0), f^3(x_0), f^4(x_0)\}$ , the following one with the five later iterations, and so on.



**Fig. 5. a)** Input to the Network



**b)** Output of the Network

Figure 5 shows the input to the network, that is, the input file consisting of 100 points from iterative Gumowski and Mira function starting from  $Y$ ; and the output of the network after the learning process. Output is a straight

line that according to the error has more or less density with more space between two points. Starting from point  $Y = (-2, 2) \in C_1$  and applying the iterative function, in a file with 100 patterns that is built with 100 iterations the obtained error is fixed to 0.00268969. But if some small random dynamic noise, uniformly distributed on interval  $[-0.1, 0.1]$ , is added to each pattern, then the error after 100 iterations is fixed to 0.00245235.

One of the more important advantages of this technique is that the controllers obtained are very stable even with small random dynamic noise or with few data.

## 2. Output Patterns

The input pattern is the equilibrium point in which the control function is going to be controlled.

When the control is made in a limit cycle, for example a limit cycle of period-5, the output layer will have five neurons that correspond to the first coordinates of the period-5 point, if the point is  $Q = (q_1, q_2)$ , the output will be constructed:  $\{q_1, f(q_1), f^2(q_1), f^3(q_1), f^4(q_1)\}$  that is the orbit of period-5.

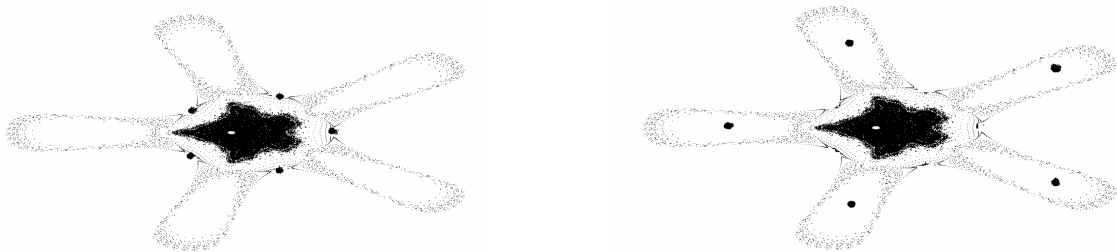


Fig. 6. 5-period orbits of the function of Gumowski and Mira located in different basin of attraction

## Estimated Model

Once the control model is designed, the following steps are taken:

1. Network weights must be fixed, always with same values in order to get a deterministic behavior.
2. An input pattern is presented and the output is calculated. To finish the learning phase of the network, another input pattern set is obtained starting from a point, finding the time series with 500 function patterns. The number of iterations to learn is 10 and the mean square error is acceptable, so the network has achieved a good solution.
3. Number of input patterns. The variation of error along the number of input patterns has been studied, among them files with different patterns. However, the error considerably decreases if the number of sweeps increases. This is due to the fact that the network is forced to learn the patterns during more iteration. Then, the error is approximately the same if the total number of patterns used to train the network agrees; therefore, if there are few data, to consider the error acceptable, it is necessary to increase the number of iterations of the network in the learning phase.

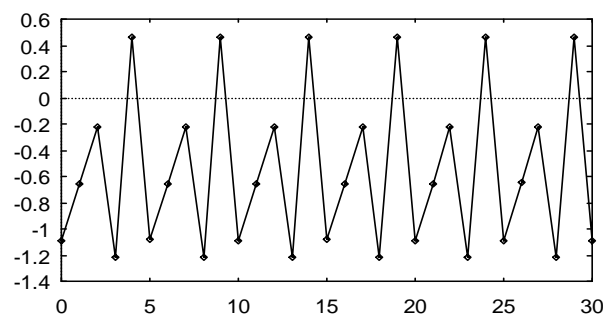


Figure.7. Real outputs of the network of the first component of the controlled Tinkerbell function around each point of the period-5 orbit

## Validate Model

Once the learning phase is completed, it is necessary to check if the network is able to control the function in the stable point.

It will be verified that the error is acceptable, because it is necessary to validate the network with different orbits and to study the outputs and the errors. It is also necessary to verify that the control model is robust; that is to say, to study the errors of the network when the patterns have some type of noise. The fastest form to verify the output errors of the network is by means of a graph.

## Applications

The control models have been designed according to the previously described procedure, based on neural networks, to control several systems in unstable periodic orbits. Discrete non-linear systems of second order in an equilibrium point have been controlled; the chosen systems are the Hénon and the Lozi [HERN99]. Another discrete system, the Ikeda, was controlled [HERN199] in a equilibrium point, and a period-5 point, that is to say, in a period-1 limit cycle, and in a period-5 limit cycle; and for the system of Tinkerbell [HERN00] in period-1, 5, and 7 limit cycles. Later, the continuous Lorenz system [HERN01] was controlled through the same procedure.

Also, the same model of control has also been applied to a dynamic system with multiple attractors and a controller has been designed to transfer and control orbits of any basin of attraction, in an equilibrium point and in a period-5 limit cycle of the different attractors of the system. The chosen system is the one of Gumowski and Mira [HERN101]. To finish, two chaotic signals were simultaneously controlled: as much for discrete as for continuous systems [HERN02], [HERN102].

## Conclusions

The main contributions of this work are:

- The construction of a controller model that uses a neural network which is very simple to design
- The control method is flexible, it can be adapted to any system, even to control and classify several systems simultaneously.
- The control can be applied in any point of the system, with no need for a waiting time before the system approaches the orbit wanted for controlling the system.
- The control is effective as long as it is applied and throughout all the time, the accumulation of computational errors does not have influence while in other methods it destabilizes the system after some iteration.
- The control is robust, his behavior is satisfactory even in the presence of random dynamic noise and although it has few data, which is very important due to the sensitivity of the initial conditions of the chaotic systems.
- The systems use the same model as much to control a system with a single one attractor in an unstable periodic orbit as to control a multi-attractor system in an unstable periodic orbit of any basin of attraction, transferring the dynamic of the system from one attractor to another.
- The model of the designed control has a great speed of learning, obtaining little errors.

The networks used for the control model have supervised learning; other types of networks could be investigated, such as associative networks; and it can also be studied the possibility of controlling chaotic systems, where the attractors are not known, training the network with orbits of the different basins of attraction.

## Bibliography

- [Barreto 1995] 'Multiparameter Control of Chaos' Ernest Barreto y Celso Grebogi, Physical Review E. Vol<sup>o</sup>2 n<sup>o</sup>4, pp 3553-3557, (1995)
- [Brandt 1995] 'Feedback Control of a Quadratic Map Model of Carciac Chaos' Michael E. Brandt y Guanrong Chen, International Journal of Bifurcation and Chaos Vol 6 n<sup>o</sup>4, pp 715-723, (1995)
- [Chen 1992] "On Feedback Control of Chaotic Dynamical Systems", G. Chen y X. Dong, Int. J.of Bifurcations and Chaos, 2, pp 407-411, (1992)



- [Garfinkel 1992] 'Controlling Cardiac Chaos', A. Garfinkel, M. L. Spano, W. L. Ditto y J. N. Weis, *cience* 257, pp 1230-1235, (1992)
- [Gulick 1992] 'Encounters with Chaos', D. Gulick, McGraw-Hill, Inc 1<sup>st</sup> edition, (1992)
- [Hübler 1989] 'Adaptive control of chaotic systems', A.W. Hübler, *Helvetica Physica* A62, pp 343-346, (1989)
- [Jackson 1991] 'Entrainment and migration controls of two-dimensions maps', E. A. Jackson and A. Kodogeorgion, *Physica* D54, pp 253-265, (1991)
- [Johnson 1995] 'Stabilized Spatiotemporal Waves in a Convectively Unstable Open Flow System: Coupled Diode Resonators', G. A. Johnson, M. Löcher y E.R. Hunt, *Phys. Rev. E* 51, pp 1625-1628, (1995)
- [Hernandez 1999] "Neural Network Control of Chaotics Systems". *Hernandez, C., Martínez, A., Castellanos, J., Computational Intelligence for Modelling, Control & Automation. Concurrent Systems Engineering Series. ISSN: 1383-7575. Vol. 54, pp. 1-8. 1999.*
- [Hernandez1 1999] "Controlling Chaotic Nonlinear Dynamical Systems". *Hernández C., Martínez A., Castellanos J., Mingo L.F.: IEEE Catalogue Number 99EX357. ISBN: 0-7802-56X2-9. Piscataway N.J. USA. pp. 1231-1234. 1999.*
- [Hernandez 2000] "Periodic Orbit Stabilization with Neural Networks". *Hernández C., Martínez A., Mingo L.F., Castellanos J.; Frontiers in Artificial Intelligence and Applications. Vol. 57. New Frontiers in Computational Intelligence and its Applications. IOS Press. ISSN: 0922-6389. pp. 109-118. 2000.*
- [Hernandez 2001] "Migration Goal Control of Chaotic Systems with Neural Networks". *Hernandez, C., Castellanos, J., Martínez, A., Mingo L.F.; Knowledge Based Intelligent Information Engineering Systems and Allied Technologies. Frontiers in Artificial Intelligence and Applications. IOS Press Ohmsha. ISSN: 0922-6389. ISBN: 1-58603-1929. Vol.: 69. Part II. pp.: 1165-1169. 2001.*
- [Hernandez1 2001] "Controlling Lorenz Chaos with Neural Networks". *Hernandez, C., Martínez, A., Mingo, L.F., Castellanos, J.; Advances in Scientific Computing, Computational Intelligence and Applications. WSES Press. ISBN: 96-8052-36-X. pp. 302-309. 2001.*
- [Hernandez 2002] "Neural Control of Simultaneous Chaotic Systems". *Hernandez, C., Gonzalo, R., Castellanos, J., Martínez, A.; Frontiers in Artificial Intelligence and Applications. Vol. 82. IOSPRESS. ISSN: 0922-6389. pp. 527-531. 2002.*
- [Hernandez1 2002] "Simultaneous Control of Chaotic Systems". *Hernández, C., Martínez, A., Castellanos, J., Luengo, C.; Recent Advances in Circuits, Systems and Signal Processing, WSEAS Press. ISBN: 960-8052-64-5. pp. 200-204. 2002.*
- [Martín 1995] 'Iniciación al caos', Miguel Ángel Martín, Manuel Morán, Miguel Reyes; Editorial Sintesis, ISB: 84-7738-293-X, (1995)
- [Nusse 1997] "Dynamics: Numerical Explorations", H.E. Nusse, J.A. Yorke, J.E. Marden y L.Sirovich (Springer-Verlag, New York). Series: Applied Mathematical Sciences V. 101, (1997)
- [Lippmann 1987] 'An Introduction to Computing with Neural Nets' L. P. Lippmann, *IEEE ASSP Magazine*, April 1987 pp. 4-22, (1987)
- [Petrov 1994] 'Controlling Chaos in the Belousov-Zhabotinsky Reaction', V. Petrov, J. Masere y K. Showalter, *Nature* 361, pp 240-243, (1994)
- [Roy 1992] 'Dynamical Control of a Chaotic Laser: Experimental Stabilization of a Globally Coupled System', R. Roy, Jr. T. W. Murphy, T. D. Maier y Z. Gills, *Phys. Rev. Lett.* 68, pp 1259-1262, (1992)
- [Sinha 1998] 'Dynamics Based Computation', Sudeshna Sinha y William L. Ditto, *Physical Review Letters* Vol 81, n°10, pp 2156-2159; (1998)
- [Sontag 1993] 'Feedback Stabilization Using Two-hidden-layer Net' Eduardo E. Sontag, *IEEE Trans, Neural Networks* 3, pp 981-990 (1993)

---

## Authors' Information

**Juan Castellanos Peñuela** – Departamento de Inteligencia Artificial, Facultad de Informática – Universidad Politécnica de Madrid (Campus de Montegancedo) – 28660 Boadilla de Monte – Madrid – Spain; e-mail: [icastellanos@fi.upm.es](mailto:icastellanos@fi.upm.es)

**Cristina Hernández de la Sota** – Departamento de Inteligencia Artificial, Facultad de Informática – Universidad Politécnica de Madrid (Campus de Montegancedo) – 28660 Boadilla de Monte – Madrid – Spain; e-mail: [cristinah@renfe.es](mailto:cristinah@renfe.es)

**Rafael Gonzalo Molina** – Departamento de Inteligencia Artificial, Facultad de Informática – Universidad Politécnica de Madrid (Campus de Montegancedo) – 28660 Boadilla de Monte – Madrid – Spain; e-mail: [rgonzalo@fi.upm.es](mailto:rgonzalo@fi.upm.es)

**Valentín Palencia Alejandro** – Departamento de Arquitectura y Tecnología de Sistemas Informáticos, Facultad de Informática – Universidad Politécnica de Madrid (Campus de Montegancedo) – 28660 Boadilla de Monte – Madrid – Spain; e-mail: [vpalencia@fi.upm.es](mailto:vpalencia@fi.upm.es)

## SYMBOLIC AND NUMERIC CONNECTIONIST MODELS SOLVING NP-PROBLEMS\*

Luis Fernando de Mingo López, Francisco Gisbert

**Abstract:** *This paper presents some connectionist models that are widely used to solve NP-problems. Most well known numeric models are Neural Networks that are able to approximate any function or classify any pattern set provided numeric information is injected into the net. Neural Nets usually have a supervised or unsupervised learning stage in order to perform desired response. Concerning symbolic information new research area has been developed, inspired by George Paun, called Membrane Systems. A step forward, in a similar Neural Network architecture, was done to obtain Networks of Evolutionary Processors (NEP). A NEP is a set of processors connected by a graph, each processor only deals with symbolic information using rules. In short, objects in processors can evolve and pass through processors until a stable configuration is reach. This paper just shows some ideas about these two models.*

**Keywords:** *Natural Computation, Membrane Systems, Neural Networks, Networks of Evolutionary Processors.*

---

### Introduction

---

Natural sciences, and especially biology, represented a rich source of modeling paradigms. Well-defined areas of artificial intelligence (genetic algorithms, neural networks), mathematics, and theoretical computer science (L systems, DNA computing) are massively influenced by the behavior of various biological entities and phenomena. In the last decades or so, new emerging fields of so-called "natural computing" identify new (unconventional) computational paradigms in different forms. There are attempts to define and investigate new mathematical or theoretical models inspired by nature, as well as investigations into defining programming paradigms that implement computational approaches suggested by biochemical phenomena. Especially since Adleman's experiment, these investigations received a new perspective. One hopes that global system-level behavior may be translated into interactions of a myriad of components with simple behavior and limited computing and communication capabilities that are able to express and solve, via various optimizations, complex problems otherwise hard to approach.

In the last decade and especially after Adleman's experiment [1] a number of computational paradigms, inspired or gleaned from biochemical phenomena, are becoming of growing interest building a wealth of models, called generically Molecular Computing. New advances in, on the one hand, molecular and theoretical biology, and on the other hand, mathematical and computational sciences promise to make it possible in the near future to have accurate systemic models of complex biological phenomena. Recent advances in cellular Biology led to new models, hierarchically organized, defining a new emergent research area called Cellular Computing.

---

### Numeric Models - Neural Networks

---

Neural networks are non-linear systems whose structure is based on principles observed in biological neuronal systems. A neural network could be seen as a system that can be able to answer a query or give an output as answer to a specific input. The in/out combination, i.e. the transfer function of the network is not programmed, but obtained through a "training" process on empiric datasets. In practice, the network learns the function that links input together with output by processing correct input/output couples. Actually, for each given input, within the learning process, the network gives a certain output that is not exactly the desired output, so the training algorithm modifies some parameters of the network in the desired direction. Hence, every time an example is input, the algorithm adjusts its network parameters to the optimal values for the given solution: in this way, the algorithm tries to reach the best solution for all the examples. These parameters we are speaking about are essentially the weights or linking factors between each neuron that forms our network. Neural Networks' application fields are typically those where classic algorithms fail because of their inflexibility (they need precise input datasets). Usually problems with imprecise input datasets are those whose number of possible input

---

\* Supported by INTAS 00-626 and TIC 2003-09319-c03-03.

datasets is so big that they can't be classified. For example in image recognition are used probabilistic algorithms whose efficiency is lower than neural networks' and whose characteristics are low flexibility and high development complexity. Another field where classic algorithms are in troubles is the analysis of those phenomena whose mathematical rules are unknown. There are indeed rather complex algorithms which can analyse these phenomena but, from comparisons on the results, it comes out that neural networks result far more efficient: these algorithms use Fourier's transform to decompose phenomena in frequential components and for this reason they result highly complex and they can only extract a limited number of harmonics generating a big number of approximations. A neural network trained with complex phenomena's data is able to estimate also frequential components, this means that it realizes in its inside a Fourier's transform even if it was not trained for that! One of the most important neural networks' applications is undoubtedly the estimation of complex phenomena such as meteorological, financial, socio-economical or urban events. Thanks to a neural network it's possible to predict, analyzing historical series of datasets just as with these systems but there is no need to restrict the problem or use Fourier's transform. A defect common to all those methods it's to restrict the problem setting certain hypothesis that can turn out to be wrong. We just have to train the neural network with historical series of data given by the phenomenon we are studying. Calibrating a neural network means to determinate the parameters of the connections (synapsis) through the training process. Once calibrated there is needed to test the network efficiency with known datasets, which has not been used in the learning process. There is a great number of Neural Networks that are substantially distinguished by: type of use, learning model (supervised/non-supervised), learning algorithm, architecture, etc.

Multilayer perceptrons (MLPs) are layered feed forward networks typically trained with static backpropagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input-output map. In principle, backpropagation provides a way to train networks with any number of hidden units arranged in any number of layers. In fact, the network does not have to be organized in layers any pattern of connectivity that permits a partial ordering of the nodes from input to output is allowed. In other words, there must be a way to order the units such that all connections go from "earlier" (closer to the input) to "later" ones (closer to the output). This is equivalent to stating that their connection pattern must not contain any cycles. Networks that respect this constraint are called feed forward networks; their connection pattern forms a directed acyclic graph or dag.

Jordan and Elman networks extend the multilayer perceptron with context units, which are processing elements that remember past activity. Context units provide the network with the ability to extract temporal information from the data. In Elman networks, the activity of the first hidden layer are copied to the context units, while the Jordan network copies the output of the network. Jordan and Elman networks combine the past values of the context unit with the present input  $x$  to obtain the present net output. The Jordan context unit acts as a so-called lowpass filter, which creates an output that is the weighted (average) value of some of its most recent past outputs.

Time lagged recurrent networks are MLPs extended with short-term memory structures. Most real-world data contains information in its time structure. Yet, most neural networks are purely static classifiers. TLRNs are the state of the art in nonlinear time series prediction, system identification and temporal pattern classification.

---

### **Symbolic Models - Cellular Computing**

---

P-systems represent a class of distributed and parallel computing devices of a biological type that was introduced in [14] which are included in the wider field of cellular computing. Several variants of this model have been investigated and the literature on the subject is now rapidly growing. The main results in this area show that P-systems are a very powerful and efficient computational model [15], [16], [13]. There are variants that might be classified according to different criteria. They may be regarded as language generators or acceptors, working with strings or multisets, developing synchronous or asynchronous computation. Two main classes of P-systems can be identified in the area of membrane computing [15]: cell-like P-systems and tissue-like P-systems. The former type is inspired by the internal organization of living cells with different compartments and membranes hierarchically arranged; formally this structure is associated with a tree. Tissue P-systems have been motivated by the structure and behavior of multicellular organisms where they form a multitude of different tissues performing various functions [2]; the structure of the system is instead represented as a graph where nodes are associated with the cells which are allowed to communicate alongside the edges of the graph.

More recently, a notion of population P-systems has been introduced [3], [4] as a model for tissue P-systems where the structure of the underlying graph can be modified during a computation by varying the set of nodes and the set of edges in the graph. Specifically, nodes are associated with cells, each of them representing a basic functional unit of the system, and edges model bonds among these cells that are dynamically created and destroyed. Although mainly inspired by the cell behavior in living tissues, population P-systems may be also regarded as an abstraction of a population of bio-entities aggregated together in a more complex bio-unit (e.g. social insects like ants, bees, wasps etc, organized in colonies or bacteria of different types). This is the main reason why we use the term population instead of tissue albeit the term cell is retained to denoting an individual in the system. The concept also recalls other similar computational models: grammar systems [8], eco-grammar systems [9], or more recently, networks of parallel/evolutionary processors [10].

Universality results have been obtained [4] for a number of variants of population P-systems. The following different rules are considered: transformation rules for modifying the objects that are present inside the cells, communication rules for moving objects from a cell to another one, cell division rules for introducing new cells in the system, cell differentiation rules for changing the types of the cells, and cell death rules for removing cells from the system. As well as this, bond-making rules are considered that are used to modify the links between the existing cells (i.e., the set of edges in the graph) at the end of each step of evolution performed by means of the aforementioned rules. In other words, a population P-system in [4] is basically defined as an evolution-communication P-system [7] but with the important difference that the structure of the system is not rigid and it is represented as an arbitrary graph. In particular, bond making rules are able to influence cell capability of moving objects from a place to another one by varying the set of edges in the underlying graph.

Another interesting variant of population P-systems is obtained by considering the general mechanism of cell communication based on signal molecules as a mechanism for triggering particular transformations inside of a cell once a particular signal-object has been received from some other cell in the system [3]. This leads to a notion of population P-systems where the sets of rules associated with the cell can vary according to the presence of particular objects inside and outside the cells. Yet again, the introduction of this mechanism is motivated by the features shared by biological systems at various levels where the behavior of an individual is affected both by its internal state and by the external stimuli received. Some results concerning the power of population P-systems with a rule activating mechanism have been obtained [5].

Further developments of the area of population P-systems are expected to cover alternative ways of defining the result of a computation and the use of string objects. Population P-systems in fact attempt to model aspects of biological systems formed by many different individual components cooperating in a coherent way for the benefit of the system as a whole; a more appropriate notion of computation is therefore necessary in order to characterize the emergent behavior of the system. Existing approaches in the area of grammar system such parallel communicating grammar systems [8] or eco-grammar systems [9], rely on the use of a single sentential form that is rewritten in parallel by different interacting/cooperating grammar components. In particular, in the case of eco-grammar systems, this sentential form is associated with the environment and it can be rewritten both by rules corresponding to action taken from the individual components in the system and by dedicated rules associated with the environment. In a similar way, we can consider string-processing population P-systems where the result of a computation is given by a string (or a language) produced in the environment at the end of a computation. However, with respect to grammar systems, population P-systems present some other interesting features like the possibility of moving objects from a place to another one, the possibility of forming bonds among the cells, the possibility of introducing new cells in the system by means of cell division, which need to be formalized for the particular case of string objects. In this respect, we aim to present some reasonable variants of population P-systems with string objects.

Membranes in P-systems can be connected using a graph and all of them can be treated as skin ones forming a so-called Network of Evolutionary Processors. A network of evolutionary processors of size  $n$  is a construct  $NEP = (V, N_1, N_2, \dots, N_n, G)$ , where  $V$  is an alphabet and for each  $1 \leq i \leq n$ ,  $N_i = (M_i, A_i, P_i, PO_i)$  is the  $i$ -th evolutionary node processor of the network. The parameters of every processor are:

- $M_i$  is a finite set of evolution rules of one of the following forms only:
  - $a \rightarrow b, a, b \in V$  (substitution rules)
  - $a \rightarrow \varepsilon, a \in V$  (deletion rules)
  - $\varepsilon \rightarrow a, a \in V$  (insertion rules)

More clearly, the set of evolution rules of any processor contains either substitution or deletion or insertion rules.

- $A_i$  is a finite set of strings over  $V$ . The set  $A_i$  is the set of initial strings in the  $i$ -th node. Actually, in what follows, we consider that each string appearing in any node at any step has an arbitrarily large number of copies in that node, so that we shall identify multisets by their supports.
- $PI_i$  and  $PO_i$  are subsets of  $V^*$  representing the input and the output filter, respectively. These filters are defined by the membership condition, namely a string  $w \in V^*$  can pass the input filter (the output filter) if  $w \in PI_i$  ( $w \in PO_i$ ).

$G = (N_1, N_2, \dots, N_n, E)$  is an undirected graph called the underlying graph of the network. The edges of  $G$ , that is the elements of  $E$ , are given in the form of sets of two nodes. The complete graph with  $n$  vertices is denoted by  $K_n$ . By a configuration (state) of an NEP as above we mean an  $n$ -tuple  $C = (L_1, L_2, \dots, L_n)$ , with  $L_i \subseteq V^*$  for all  $1 \leq i \leq n$ . A configuration represents the sets of strings (remember that each string appears in an arbitrarily large number of copies) which are present in any node at a given moment; clearly the initial configuration of the network is  $C_0 = (A_1, A_2, \dots, A_n)$ .

A configuration can change either by an evolutionary step or by a communicating step. When changing by an evolutionary step, each component  $L_i$  of the configuration is changed in accordance with the evolutionary rules associated with the node  $i$ . When changing by a communication step, each node processor  $N_i$  sends all copies of the strings it has which are able to pass its output filter to all the node processors connected to  $N_i$  and receives all copies of the strings sent by any node processor connected with  $N_i$  providing that they can pass its input filter.

**Theorem 1.** Each recursively enumerable language can be generated by a complete NEP of size 5. [21]

**Theorem 2.** Each recursively enumerable language can be generated by a star NEP of size 5. [22]

**Theorem 3.** The bounded PCP can be solved by a NEP in size and time linearly bounded by the product of  $K$  and the length of the longest string of the two Post lists. [23]

A simple NEP of size  $n$  is a construct  $SNEP = (V, N_1, N_2, \dots, N_n, G)$ , where,  $V$  and  $G$  have the same interpretation as for NEPs, and for each  $1 \leq i \leq n$ ,  $N_i = (M_i, A_i, PI_i, FI_i, PO_i, FO_i)$  is the  $i$ -th evolutionary node processor of the network.  $M_i$  and  $A_i$  from above have the same interpretation as for an evolutionary node in a NEP, but:

- $PI_i$  and  $FI_i$  are subsets of  $V$  representing the input filter. This filter, as well as the output filter, is defined by random context conditions,  $PI_i$  forms the permitting context condition and  $FI_i$  forms the forbidding context condition. A string  $w \in V^*$  can pass the input filter of the node processor  $i$ , if  $w$  contains each element of  $PI_i$  but no element of  $FI_i$ . Note that any of the random context conditions may be empty, in this case the corresponding context check is omitted. We write  $\rho_i(w) = \text{true}$ , if  $w$  can pass the input filter of the node processor  $i$  and  $\rho_i(w) = \text{false}$ , otherwise.
- $PO_i$  and  $FO_i$  are subsets of  $V$  representing the output filter. Analogously, a string can pass the output filter of a node processor if it satisfies the random context conditions associated with that node. Similarly, we write  $\tau_i(w) = \text{true}$ , if  $w$  can pass the input filter of the node processor  $i$  and  $\tau_i(w) = \text{false}$ , otherwise.

**Theorem 4.** The families of regular and context-free languages are incomparable with the family of languages generated by simple NEPs. [21]

**Theorem 5.** The "3-colorability problem" can be solved in  $O(m + n)$  time by a complete simple NEP of size  $7m+2$ , where  $n$  is the number of vertices and  $m$  is the number of edges of the input graph. [21]

---

## Conclusions

This paper has introduced the novel computational paradigm Networks of Evolutionary Processors. Connectionists' models such as Neural Networks can be taken into account to develop NEP architecture in order to improve behavior. As a future research, learning concepts in neural networks can be adapted in a NEP architecture provided the numeric-symbolic difference in both models.

---

**Bibliography**


---

- [1] Adleman, L.M. 1994. Molecular computation of solutions to combinatorial problems. *Science*, 226, 1021-1024
- [2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. *The Molecular Biology of the Cell*. Fourth Edition. Garland Publ. Inc., London
- [3] Bernardini, F., Gheorghe, M. 2004. Cell Communication in Tissue Psystems and Cell Division in Population P-Systems. In [17], 74-91
- [4] Bernardini, F., Gheorghe, M. 2004. Population P-Systems. *Journal of Universal Computer Science*, 10, 509-539
- [5] Bernardini, F., Gheorghe, M. 2005. Cell Communication in Tissue PSystems: Universality Results. *Soft Computing* (to appear)
- [6] Busby, S., de Lorenzo, V. 2001. Cell regulation - putting together pieces of the big puzzle. *Curr. Op. Microbiol*, 4, 117-118
- [7] Cavaliere, M. 2003. Evolution Communication P-Systems. In [16], 134-145 and in [17], 206-223
- [8] Cshaj-Varju, E., Dassow, J., Kelemen, J., Paun, Gh. 1997. *Grammar Systems. A Grammatical Approach to Distribution and Cooperation*. Gordon and Breach, London
- [9] Cshaj-Varju, E., Kelemen, J., Kelemenova, A., Paun, Gh. 1997. *EcoGrammar Systems: A Grammatical Framework for Studying Life-Like Interactions*. *Artificial Life*, 3, 1-28
- [10] Cshaj-Varju, E., Salomaa, A., 1997. Networks of Parallel Language Processors. In *New Trends in Formal Languages. Control, Cooperation, and Combinatorics*. In P. Paun, Gh., Salomaa, A. (eds), *Lecture Notes in Computer Science*, 1218, Springer-Verlag, Berlin, Heidelberg, New York, 299-318
- [11] Krasnogor, N., Gheorghe, M., Terrazas, G., Diggie, S., Williams, P., Camara, M. 2005. *Bulletin of the EATCS* (to appear)
- [12] Jennings, N.R. 2000. On agent-based software engineering. *Artificial Intelligence*, 117, 277-296
- [13] Martin-Vide, C., Mauri, G., Paun, Gh., Rozenberg, G., Salomaa, A. (eds) 2004. *Membrane computing. International workshop, WMC 2003, Tarragona, Spain, July 2003. Revised papers*. *Lecture Notes in Computer Science*, 2933, Springer, Berlin Heidelberg New York
- [14] Paun, Gh. 2000. Computing with Membranes. *Journal of Computer and System Sciences*, 61, 108-143
- [15] Paun, Gh. 2002. *Membrane computing. An introduction*. Springer, Berlin Heidelberg New York
- [16] Paun, Gh., Rozenberg, G., Salomaa, A., Zandron, C. (eds) 2003. *Membrane computing. International workshop, WMC-CdeA 02, Curtea de Arges, Romania, August 19-23, 2002. Revised papers*. *Lecture Notes in Computer Science*, 2597, Springer, Berlin Heidelberg New York
- [17] Paun, Gh., Riscos-Nunez, A., Romero-Jimenez, A., Sancho-Caparrini, F. (eds) 2004. *Second brainstorming week on membrane computing, Seville, 2-7 February 2004. Technical Report 01/2004, Department of Computer Science and Artificial Intelligence, University of Seville, Spain*
- [18] Swift, S., Downie, J.A., Whitehead, N.A., Barnard, A.M.L., Salmond, G.P.C., Williams, P. 2001. Quorum sensing as a population-density-dependent determinant of bacterial physiology. *Adv Micro Physiol*, 45, 199-270
- [19] Williams, P., Camara, M., Hardman, A., Swift, S., Milton, D., Hope, V.J., Winzer, K., Middleton, B., Pritchard, D.I., Bycroft, B.W. 2000. Quorum sensing and the population dependent control of virulence. *Phil. Trans Roy Soc London B*, 355(1397), 667-680
- [20] Winzer, K., Hardie, K.H., Williams, P. 2002. Bacterial cell-to-cell communication: sorry can't talk now – gone to lunch!. *Curr Op. Microbiol*, 5, 216-222
- [21] Juan Castellanos, Carlos Martin-Vide, Victor Mitrana, Jose M. Sempere; *Networks of Evolutionary Processors*
- [22] Juan Castellanos, Carlos Martin-Vide, Victor Mitrana, Jose M. Sempere; *Solving NP-Complete Problems with Networks of Evolutionary Processors*. IWANN
- [23] Carlos Martin-Vide, Victor Mitrana, Mario J. Perez-Jimenez, Fernando Sancho-Caparrini; *Hybrid Networks of Evolutionary Processors*. *GECCO 2003. Lecture Notes on Computer Science* 2723, pp. 401-412. 2003

---

**Authors' Information**


---

**Luis Fernando de Mingo López** – Dept. Organización y Estructura de la Información, Escuela Universitaria de Informática, Universidad Politécnica de Madrid, Crta. De Valencia km. 7, 28031 Madrid, Spain; e-mail: [lfmingo@eui.upm.es](mailto:lfmingo@eui.upm.es)

**Francisco Gisbert** – Dept. Lenguajes y Sistemas Informáticos e Ingeniería del Software, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain; e-mail: [francisco.gisbert@upm.es](mailto:francisco.gisbert@upm.es)

---

## FROM TEXTUAL TO COMPUTATIONAL INFORMATION MODELLING

**Jesús Cardeñosa, Carolina Gallardo, Eugenio Santos**

**Abstract:** *Information can be expressed in many ways according to the different capacities of humans to perceive it. Current systems deals with multimedia, multiformat and multiplatform systems but another « multi » is still pending to guarantee global access to information, that is, multilinguality. Different languages imply different replications of the systems according to the language in question. No solutions appear to represent the bridge between the human representation (natural language) and a system-oriented representation. The United Nations University defined in 1997 a language to be the support of effective multilinguism in Internet. In this paper, we describe this language and its possible applications beyond multilingual services as the possible future standard for different language independent applications*

**Keywords:** *Knowledge Representation, Information modelling*

---

### Introduction

The concept of Multimedia systems could be broadly defined as the set of systems built for transmitting information through any means that human beings are able to catch. People communicate in a natural way through the five senses, either individually or cooperatively. The evolution of technology has permitted the existence of systems in the market that reproduce the natural five senses. Among these systems and technologies, the most classical and frequent systems to transmit information utilize sight and hearing. These natural senses have permitted to see and hear. They have allowed us to visualize images of all kinds, from natural images to drawings and pictures, and from static to animated images. Further, we have been able to perceive them combined with natural or artificial sounds. However, sight and hearing do not constitute the most massive means to transmit information; the most appealed media to transmit knowledge among human beings is language, mainly recorded in its written form.

Ever since immemorial times, when human beings wanted knowledge to endure, they recorded in written texts. Even today, in the image era, images are accompanied by text. The profusion of support media for information has made that classical systems based on natural language texts incapable of organizing the information in an adequate manner, hindering a correct management of information. Several technologies like document management or information retrieval have been helped by great computational systems in order to palliate such situations.

From the second half of the XX century, when the production of written information has been massive, Internet has put in an appearance, and worldwide commerce has grown exponentially. Media companies have started to commercialize information by itself, which means that they have to conceive systems for storing information in an organized and more compact way, easier to find and thus easier to be offered to those requesting it. This requires of complex systems but also of ways of representing knowledge that permits the reliable interchange of information between humans and machines. Human language is the externalization of human knowledge and the vehicle for knowledge exchange among humans but it is inadequate for machines.

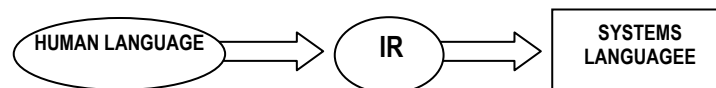
An intelligent use of knowledge for any purpose (like question answering, information retrieval, concepts deduction, etc.) calls for mechanisms of representation completely devoid of any source of ambiguity and imprecision found in natural languages, while endowed with the formal properties characteristic of computational languages.

For that, we should invest more resources in the processes of capturing, organizing and diffusing information. Thus, in the same way that we associate *knowledge* to the “sources”, we should ascribe *information* to “systems”, i.e., to the media in charge of providing information. That is, *knowledge* and *information* is not the same thing, and the transformation of one into the other entails something more than human language. It entails a system that permits the conversion of one into the other.

However, the design of an intermediate system between human language, as a way to represent knowledge, and systems language is not a trivial issue. In fact, human language is self-organized in an unconscious manner; it represents highly complex mental processes (be it in written or oral form) such as searching and rearranging data in a way that can be useful and comprehensible for others. On the other hand, systems for the capture and diffusion of information have a conscious organization but an information search capability quite limited. That is:

	Knowledge Organization	Functionalities
<b>Human language</b>	Unconscious	High
<b>Systems language</b>	Conscious	Low

The gap between the organization of human knowledge and systems knowledge still remains as the main obstacle for the construction of efficient systems with conscious knowledge, able to be easily maintained, modified or expanded. A possible approach to “fill this gap” could be an intermediate representation of knowledge (hereafter *IR*) serving as a bridge between human language and systems language. That is, a representation able to express contents with schemata and models close to human beings but at the same time a representation able to communicate with systems or represent knowledge at the systems level. Figure 1 illustrates the intermediary role that the IR language could play among human languages and systems language.



**Figure 1.** Intermediate Language between human language and systems language

There have already been attempts to develop these languages, although not associated to the use that we propose here. We are referring to Schank’s theory of Conceptual Dependencies [Schank, 1972] to other models such as Sowa’s [Sowa, 1996] that have been in fact predecessors of current ontologies. However, these models lacked a number of applications and a wider computational capability when they were proposed. In another sector of the industry, concretely language industry, we can find similar models used to represent contents in a language-independently manner. These models were known as “Interlinguas” and were used in machine translation systems such as PIVOT [Muraki, 1989] and ATLAS [Uchida, 1989].

We could say that, although with a different perspective, there are already precedents of languages and models that have attempted to overcome the gap between human languages and machine languages. We will generically call *Intermediate Representation Language* to the knowledge representation language able to make compatible human languages and systems languages.

In the late 90ies, the University of the United Nations started the development of a computational language that would allow for the representation of human knowledge in a language-independent way. The rationale behind this was to the elimination of linguistic barriers for content access in Internet. The mere fact of guaranteeing language independence involved the representation of the deepest structures of conscious human expression, so that written texts in any natural language would have the same representation in that computational language. After years of research and testing, such language has been reaffirmed as a language able to be understood by systems (for example, in order to generate any other human natural language) and it is possible to generate it just following its specifications and corresponding manuals [Uchida, 2003]).

This approach is resulting to be much more important that the initially posed (pivotal representation of texts for the elimination of linguistic barriers). This language has been gradually transformed into a knowledge representation language starting from written texts.

In this article it is presented a description of this language and how from written texts we can achieve a representation that not only “stores” knowledge but allows for an intelligent use of the resulting representation for different purposes, provided a number of functions.



---

## The Language

---

UNL is an artificial language designed for unambiguously expressing the informational content of natural language texts in a language-independent way, with the main aim of facilitating automatic multilingual information exchange on the web or in other local contexts.

Information encoded in UNL is organised into UNL documents. Since documents are commonly organised themselves into paragraphs and these into sentences, a UNL document mimics such structure and is organised into UNL paragraphs and sentences by means of HTML-like tags. UNL paragraphs consist of UNL sentences, which in turn are divided into sections. Each UNL sentence is divided into the following sections:

- The original sentence, i.e. the information that has been encoded.
- The UNL code corresponding to the original sentence.
- For each language for which a UNL Generator has been developed the automatically generated text of the UNL code into that language. Generation results are then cached in the document and available to the reader without delay. Of course, the stored results can be renewed as soon as the generators improve their output.

Besides these elements, a UNL header contains information and meta-information about the document as a whole. Although not devised in principle, UNL documents also allow for its integration into XML, since document structuring such as tables, sections, subsections, etc are not easily expressed by the UNL machinery. In fact, it is not the objective of UNL to provide document structuring but to provide a semantic structuring of contents. These ideas are further explored in [Cardeñosa, 2005] and [Hailaoui, 2005].

A UNL expression takes the form of a directed hyper-graph. Its simple nodes contain the so-called *Universal Words* and its arcs are labelled with *conceptual relations*. In addition to simple nodes, hyper-nodes are also allowed as origin or destination of arcs and consist on UNL graphs themselves. In addition to universal words and conceptual relations, *attributes* are the third ingredient of a UNL hyper-graph. Attributes may occur as labels of the universal words, modifying them in certain key aspects. These are the three building blocks of the inter-lingua and we now turn to describe each one in detail.

### Universal Words

Universal words are so called because they attempt to be universally applicable to any natural language and because their meaning is derived from the meaning of natural language words. UWs are based on the English headwords. Initially any English headword is a candidate for becoming an UW, being its meaning the meaning defined in any authoritative monolingual dictionary of English. However, in deciding on English headwords as the building blocks of the vocabulary of the interlingua, two key aspects have to be resolved:

- a) Inherent ambiguity in English headwords.
- b) Mismatch among lexicalized concepts in English and lexicalized concepts in other natural languages.

### *Inherent Ambiguity in English Headwords*

Most natural language words are subject to ambiguity and polysemy. A single word in a natural language contains several senses (often related, often not), so it is fairly rare that the relation between a concept and a word is one-to-one. English vocabulary is not devoid of such ambiguity and thus a system based on English headwords as interlingual concepts should establish mechanisms for reducing such ambiguity. In order to reduce ambiguity, UWs are modified by semantic restrictions. Such semantic restrictions try to select a given sense or concept of an English Headword from the others. The most basic and simple way to achieve this is by attaching to the word a hypernym.

We will illustrate the process of defining UWs taking the following sentence as input:

```
Member States should, whenever necessary or desirable, conclude bilateral
agreements to deal with matters of common interest arising out of the
application of the present Recommendation.
```

Only content words (mainly nouns, verbs, adjectives and some prepositions and adverbs) require an UW, in this sentence, the first content word is *State*, which is a highly ambiguous headword in English.

*State* can be both a noun and a verb in English. Initially, there are there are two obvious candidates for “state” (being “icl” the abbreviation for “is included in” or the traditional “is a” relation) using the most general hypernym as semantic restriction:

*statet(icl>thing)* → for the nominal senses  
*state(icl>do)* → for the verbal senses

However these restrictions are not very informative and there is still a great degree of ambiguity in each of these potential UWs. In order to overcome this ambiguity, more and finer semantic restrictions have to be attached to the basic UWs. For this, the possibilities are the following:

- |   |   |
|---|---|
| <ol style="list-style-type: none"> <li>1. Use of a closer hypernym.             <ol style="list-style-type: none"> <li>a. <i>state(icl&gt;government)</i></li> <li>b. <i>state(icl&gt;region)</i></li> <li>c. <i>state(icl&gt;circumstance {&gt;abstract thing})</i></li> </ol> </li> </ol> | <ol style="list-style-type: none"> <li>2. Use of argument structure (for verbal senses):             <ol style="list-style-type: none"> <li>a. <i>state( {icl&gt;do} agt&gt;thing, obj&gt;thing)</i></li> </ol> </li> </ol> |
|---|---|

### **Lexical Mismatches among Languages**

This is the second problem that in fact, every *IL* has to tackle, and in the context of UNL, the solution does not come at first sight. Lexical mismatches arise when:

- a) For a given concept in a given language, there is not English headword. This is a frequent case for cultural-specific terms (and other not so cultural-specific).
- b) For a given concept in English, there is not an appropriate term in the target natural language. Example: En. *misunderstand* → Sp. *entender mal*
- c) There is not a one-to-one relation among languages, an English headword is covered in the target language by more than one headword:
  - a. *Corner* → Sp. *esquina* & *rincón*
  - b. *Marry* → Ru. *zhenit'sja* & *vyxodit' zamuzh*

For these three situations, a solution must be provided. So, the solution adopted in each case is the following:

- a) For the lexical gap in the English language, the specifications of the interlingua propose to include the original word as the basic UW and then semantic restrictions would be added to describe the intended meaning as much as possible, like in the Japanese term *Ikebana(icl>flower arrangement)*.
- b) Lexical gaps in the target languages can be considered as a “local” problem, to be treated in the dictionaries of target languages, and not in the design of the Interlingua. For example, in the case of “misunderstand” and Spanish, it will be the task of Spanish developers’ dictionary to link such a word with a complex expression such as “*entender mal*”.
- c) When an English headword combines the meaning of more than one headword in another language, it is possible to appeal to the semantic restrictions again in order to clarify the intended concept. In fact there are two possibilities:
  - a. Ignore the difference, the Anglo-centered vocabulary here is imposed, and it will be the task of local generator to choose (in the case of Russian) into the verb *vyxodit' zamuzh* or the verb *zhenit'sja*.
  - b. Express the difference, with the use of the semantic restrictions like for example:
    - i. [*vyxodit' zamuzh*] *marry(agt>female)*; [*zhenit'sja*] *marry(agt>male)*
    - ii. [*esquina*] *corner(mod>outside angle)*; [*rincón*] *corner(mod>inside)*

Of course it will be desirable that all these UWs conforms a hierarchy of inter-related UWs, so that *marry(agt>female)* and *marry(agt>male)* depend of a more general UW like *marry(agt>person, obj>person)*. Thus, semantic restrictions impose themselves a hierarchy into the system of UWs. The result is the so-called UNL-KB organizing the UW concepts à la *Wordnet* [Fellbaum, 1998], thus implicitly linking and relating the vocabulary of natural languages through the pivotal UW system.

## UNL Relations

The second ingredient of UNL is a set of conceptual relations. Relations form a closed set defined in the specifications of the inter-lingua. The rationale behind conceptual relations is twofold:

1. To characterize a set of semantic notions applicable to most of the existing natural languages. For instance, the notion of initiator or cause of an event (agent) is considered one of such notions since it is found in most languages.
2. To select a small set of semantic notions relevant to produce an inter-lingual semantic analysis. The notion of agent is regarded as one of such relations, because of its central role in the analysis of the meaning of many sentences.

Therefore, a UNL representation is mainly based on a role-based description of an event or situation, following the tradition started by Fillmore's case grammars, rather than a more logical semantic analysis of sentences. When defining the intended meaning of each conceptual relation, the specifications of the language [Uchida, 2003] rely on two intensional expedients:

1. Setting semantic constraints over the universal words that can appear as first (origin) and second argument (destination) of the relation. These constraints are based on the lexical relations established among universal words in the Knowledge Base. In the case of the agent relation, such restrictions include that the origin Universal Word must denote an event that accepts an initiator (and not an event that "just happens") and the destination universal word must denote an entity (as opposed to a quality or an event, let us say).
2. Giving a natural language explanation of the intended meaning of the relation. For the agent relation this explanation may just say that the agent is the cause or initiator of the event, an entity without which the event would not happen.

Provided that the Knowledge Base is built upon unambiguously defined principles, the first mechanism gives us a rigorous characterization of the conceptual relations. The second expedient is subject to the ambiguities of natural language, and the resulting definition is therefore semi-formal.

The current specification of UNL includes 41 conceptual relations, including causal, temporal, logical, numeric, circumstantial and argument relations.

Selecting the appropriate conceptual relation plus adequate universal words allows UNL to express the propositional content of any sentence.

In the input sentence, some of these relations are exemplified. This sentence describes a main event, denoted by the English predicate *conclude* and its dependent participants. The UW for the main predicate of the sentence is *conclude(agt>thing, obj>agreement)*. It requires two participants:

- a) The agent or initiator of the event: In this sentence, the agent is "Member States" that coincides with the subject of the clause.
- b) An object (or theme affected by the event, required for the completion of the event), realized in the "bilateral agreement" as the direct object.

Given the syntax of UNL and the binary nature of relations, when specifying the UNL representation of the sole event "Member States should conclude bilateral agreements" the *agt* relation links *conclude(agt>thing, obj>agreement)* as source UW and *state(icl>government)* as target UW. Analogously, there is a modifying *mod* relation between *state* and *member*, and *agreement* and *bilateral*. Figure 2 shows a graphical version of the UNL representation of the main clause of the sentence.

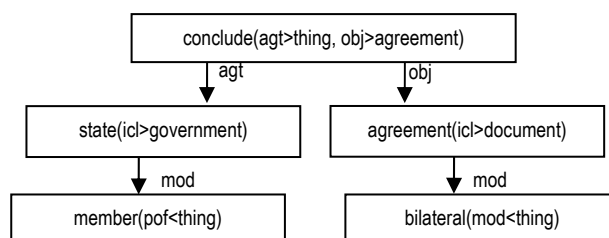


Fig. 2. Representation of the main event of the sentence

---

### Description of Attributes

Contextual information such as the time of the event with respect to the time of the utterance, informative structure of the sentence, speaker's communicative goal and attitudes, etc. is expressed in UNL by means of *attribute labels*. UNL attributes include notions such as:

- Information depending on the speaker, such as time of the described event with respect to the moment of the utterance; the communicative goal of the utterance; epistemic and deontic modality.
- Contextual information affecting both the participants both the predicate of the sentence, like aspectual properties of the event, number of nominal concepts.
- Pragmatic notions like the organization of the information in the original sentence, referential status of referring expressions and other labels determining discourse structure.
- Typographical and orthographical conventions. These include formatting attributes such as *double quotations*, *parenthesis*, *square brackets*, etc.

---

### Applications

The UNL System has an indubitable application to all the existing information systems. The introduction of multilinguality in any other system is almost a model case of added value services. However, it does not mean that multilinguality be only translation services. We will describe some possible applications of the UNL system.

#### UNL as Language for Knowledge Representation

UNL is mainly used as a support language for multilingual generation of contents coming from different languages. However, its design allows for non-language centered applications, that is, UNL could serve as a support for knowledge representation in generic domains. When there is a need to construct domain-independent ontologies, researches turn back to natural language (such as Wordnet, GUM [Bateman, 1995] or even CyC [www.cyc.com]) to explore the "semantic atoms" that knowledge expressed in natural languages is composed of. UNL follows this philosophy, since it provides an interlingual analysis of natural language semantics. The reasons why UNL could be backed as a firm knowledge representation language are:

1. The set of necessary relations existing between concepts is already standardized and well defined.
2. It is the product of intensive research on the thematic roles existing in natural languages by a number of experts in the area of Machine Translation and Artificial Intelligence, guaranteeing wide coverage of all contents expressed in any natural language.
3. Similarly, the set of necessary attributes that modify concepts and relations is fixed and well defined, guaranteeing a precise definition of contextual information.
4. UNL syntax and semantics are formally defined.

However, to really serve as a language for knowledge representation, it must support deduction mechanisms and must specify how a knowledge base could be build up in the UNL language. This idea is explored in [Cardeñosa, 2004].

Being a language suitable for knowledge representation, UNL could be the support of ontologies or of Cross Language Information Retrieval systems. UNL could be a firm candidate for this because of its long history as an interlingua, and the existence of analysis and generation systems to and from UNL.

#### Cross-lingual Information Retrieval

To support cross-lingual information search could be one of the most appealing applications. Because the information existing in UNL is in fact independent of the original language, placing information in a web site written in UNL supposes that is accessible from any other language. But also, the search systems could try to find information based on concepts (much more effective than based on terms or keywords that are dependent on the language) and find it (if it exists) independently of the language used by the generator of the information searched. UNL offers a promising approach to this kind of systems because the search of information based on concepts is not difficult to be re-written in UNL (they could be UW) almost automatically, always under the supervision of the searcher.

### **Multilingual Information System**

One possible situation is that an organization has public information that should be shared and distributed in multilingual form. This is not exactly a problem of translators (at an acceptable cost) but a problem of maintenance. In this case, an organism, such as any derived from the United Nations for instance (or any other as the European Commission, Health Care Organisation, etc.) should maintain an on-line system with all kind of information about organizations. This could be classified as a simple multilingual service, where the information should be written in UNL (in the UNL Document Base) and shown through the Web in different languages. The maintenance is carried out by making changes in the UNL code. The style in which organisations are described makes post-editions unnecessary most of the times.

An additional use of this kind of system is the maintenance of technical documentary databases with multilingual necessities. One case would be the technical documentation (maintenance of industrial equipment for instance) that has to be managed in many countries, or in the case of companies with branches in several countries where a clear and precise documentation is essential for reaching a leading market position. Writing this technical documentation in UNL would clearly permit unified contents so that no differences derived from different translators could cause technical problems. It is well known that the manuals for some languages are almost in all cases translated from the original language into English and from English to the target language introducing in some cases a double risk of mistakes. The use of the UNL system for this kind of application permits also that the post-edition (if needed) can be done directly by the final users.

The multilingual Access to Public Sector Information is a general goal of big public organizations. One of the major problems to reach the objective to make the public information really available is the multilingual origin of the documents. In fact one of the recommended actions mainly to the European Industry is to affiliate contents to the Multilingual e-Content Europe portal [Nicholas, 2000]. Of course, having multilingual contents, the industry and public organizations have also to guarantee the multilingual access.

### **Multilingual Transactional System**

Different issues have been solved in the last years to facilitate the Business to Customer (B2C) services as the typical practice of Electronic Commerce. Most of the systems are based on the use of English language. The incorporation of multilingual capacities in companies supposes an effective increasing of market and also of image. The advantage is that the amount of work to encode any text that belongs to a web-site is the same disregarding the number of source languages. It is only needed the target language generator (at the moment there exist more than ten languages generators covering the 85% of the human population). The integration of the systems is very easy and has not special complications.

However, where the UNL system should have more impact is on the B2B activities. All the concepts and components from the B2B, but particularly the ontologies based on cXML [Merkov, 1999], to define business documents defining technical and business dictionaries (completely compatible with the UW dictionaries). OBI's [OBI, 1999] data formats rely on EDI standards for document exchanges etc. which are completely compatible with the structure of the UNL Documents. The international commercial exchange and current growth of the E-commerce are due to this kind of exchanges. Here the availability of multilingual systems able to support the exchange of documents, transactional information, a correct common understanding of contractual documentation (well addressed by a common UNL codification) is perhaps the most important application of this system. In addition, corporate information and even more complex systems (as multilingual e-mail) can be supported.

### **From Bilingual to Multilingual Translation Systems**

The UNL system could be viewed as an alternative to the classical machine translations systems. However, it is not exactly the case. When the classical machine translation systems massively follow the model of "transfer", the UNL is conceived in a different way. First of all, the UNL system is not a system to support machine translation but multilingual services. It is not the same. There is not any automatic conversion from a language to UNL.

For instance, analysers and dictionaries of a particular language can be integrated with the production of UNL code at the required level. An existing dictionary can be reused to develop the UW and thus to develop the UNL Dictionary for a language or a specific domain. The target language generators of an existing language can be

reused once integrated the input with the UNL code. These operations permit the transformation of a bilingual system into a multilingual one. In fact, the Russian Language Centre\* [<http://www.unl.ru>], the French Language Centre and some others are sustaining the UNL system reusing bilingual pre-existing machine translation systems. Thus, this means that there are two types of users of this system, the industry itself manufacturing the integration and therefore creating multilingual systems with a high degree of reuse of the linguistic repository and the final users of the machine translation systems like human translators, that are normally in charge of the post-editing of the target documents. The main advantage is that these persons will increase their productivity because while working in just one language, they are producing contents in many other languages.

---

## References

- [Bateman, 1995] Bateman, J.A; Henschel, R. and Rinaldi, F. The Generalized Upper Model 2.0. 1995. Available on line at <http://www.darmstadt.gmd.de/publish/komet/gen-um/newUM.html>
- [Cardeñosa, 2004] Cardeñosa, J., Gallardo, C., and Iraola. "The forgotten key point for assuring knowledge consistency in CLIR systems". In: Proceedings of the Workshop Lessons Learned from Evaluation: Towards Transparency & Integration in Cross-Lingual Information Retrieval (LREC, 2004), Lisbon, May, 2004.
- [Cardeñosa, 2005]. Cardeñosa, J., Gallardo, C., and Iraola, L. *An XML-UNL Model for Knowledge-Based Annotation*. Research on Computing Science, vol 12, pp 300-308. ISBN: 970-36-0226-6. México, 2005
- [CyC] [www.cyc.com](http://www.cyc.com)
- [Fellbaum, 1998] Fellbaum, C., editor. WordNet: An Electronic Lexical Database. Language, Speech, and Communication Series. MIT Press, 1998
- [Hailaoui, 2005]. Hailaoui, N. and Boitet, C. A Pivot XML-Based Architecture for Multilingual, Multiversion Documents: parallel monolingual documents aligned through a central correspondence descriptor and possible use of UNL. Research on Computing Science, vol 12, pp 309-326. ISBN: 970-36-0226-6. México, 2005
- [Merkov, 1999] Merkov, M. cXML: A new Taxonomy for E-Commerce. 1999. Available on line at [http://ecommerce.internet.com/outlook/article/0,1467,7761\\_124921,00.html](http://ecommerce.internet.com/outlook/article/0,1467,7761_124921,00.html)
- [Muraki, 1989] Muraki, K. PIVOT: Two-phase machine translation system. Proceedings of the Second Machine Translation Summit, Tokyo, 1989
- [Nicholas, 2000] Nicholas, L.; and Lockwood, R. Final Report: SPICE-PREPII: Export potential and linguistic customization of digital products and services. EPS Ltd and Equipe Consortium Ltd, 2000
- [OBI, 1999] The Open Buying on the Internet (OBI) Consortium. 1999. Open buying on the Internet. OBI specification. 2000. Available on line at <http://www.openbuy.org/>
- [Schank, 1972] Schank, R.C. Conceptual Dependency: A Theory of Natural Language Understanding, Cognitive Psychology, Vol 3, 532-631, 1972
- [Sowa, 1996] Sowa, J.F. "Processes and participants," in Eklund et al., eds. (1996) Conceptual Structures: Knowledge Representation as Interlingua, Lecture Notes in AI -1115, Springer-Verlag, Berlin, pp. 1-22, 1996
- [Uchida, 1989] Uchida, H. ATLAS-II: A machine translation system using conceptual structure as an Interlingua. Proceedings of the Second Machine Translation Summit, Tokyo, 1989
- [Uchida, 2003] Uchida, H. The Universal Networking Language. Specifications. 2003. Available on-line at <http://www.unl.org>

---

## Authors' Information

**Jesús Cardeñosa** – UPM.Artificial Intelligence Department, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n. 28660 Madrid, SPAIN; email: [carde@opera.dia.fi.upm.es](mailto:carde@opera.dia.fi.upm.es)

**Carolina Gallardo** – UPM. Artificial Intelligence Department, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n. 28660 Madrid, SPAIN; email: [carolina@opera.dia.fi.upm.es](mailto:carolina@opera.dia.fi.upm.es)

**Eugenio Santos** – UPM. Business organisation Department. Universidad Politécnica de Madrid. Carretera de Valencia Km.7. 28041 Madrid, SPAIN; email: [esantos@eui.upm.es](mailto:esantos@eui.upm.es)

---

\* Language Centres are the operational national units that support the development of specific local language within the worldwide organization of the UNL Programme.

## A NEW APPROACH FOR ELIMINATING THE SPURIOUS STATES IN RECURRENT NEURAL NETWORKS

**Víctor Giménez-Martínez, Carmen Torres,  
José Joaquín Erviti Anaut, Mercedes Perez-Castellanos**

**Abstract:** As is well known, the Convergence Theorem for the Recurrent Neural Networks, is based in Lyapunov's second method, which states that associated to any one given net state, there always exist a real number, in other words an element of the one dimensional Euclidean Space  $\mathbb{R}$ , in such a way that when the state of the net changes then its associated real number decreases. In this paper we will introduce the two dimensional Euclidean space  $\mathbb{R}^2$ , as the space associated to the net, and we will define a pair of real numbers  $(x, y)$ , associated to any one given state of the net. We will prove that when the net change its state, then the product  $x \cdot y$  will decrease. All the states whose projection over the energy field are placed on the same hyperbolic surface, will be considered as points with the same energy level. On the other hand we will prove that if the states are classified attended to their distances to the zero vector, only one pattern in each one of the different classes may be at the same energy level. The retrieving procedure is analyzed trough the projection of the states on that plane. The geometrical properties of the synaptic matrix  $W$  may be used for classifying the  $n$ -dimensional state-vector space in  $n$  classes. A pattern to be recognized is seen as a point belonging to one of these classes, and depending on the class the pattern to be retrieved belongs, different weight parameters are used. The capacity of the net is improved and the spurious states are reduced. In order to clarify and corroborate the theoretical results, together with the formal theory, an application is presented

**Keywords:** Learning Systems, Pattern Recognition, Graph Theory, Image Processing, Recurrent Neural Networks.

### 1. Introduction

The problem to be considered when Recurrent Neural Networks (RNN) are going to be used as *Pattern Recognition* systems, is how to impose prescribed prototype vectors  $\xi^1, \xi^2, \dots, \xi^p$ , of the space  $\{-1, 1\}^n$ , as fixed points. In the classical approach, the synaptic matrix  $W = (w_{ij})$  should be interpreted as a sort of sign correlation matrix of the prototypes. The element  $w_{ij} \in W$ , is going to represent some kind of relation between coincidences and not coincidences on the list of the components "i" and "j" for all the prototype vectors  $\xi^1, \xi^2, \dots, \xi^p$ . The classical solution to impose fixed points by means of the synaptic matrix  $W$  is the *Hebb's* law, which states that the synaptic weight  $w_{ij}$  should increase whenever neurons "i" and "j" have simultaneously the same activity level and it should decrease in the opposite case. As it was pointed out above, the prototype vector components must belong to the set  $\{-1, 1\}$ ; this fact is the cornerstone of the *Hebb's* law mathematical interpretation. The reason is that when the prototype  $\xi^\mu$  is stored, neurons "i" and "j" may receive a similar sign or not". The mathematical advantage of this interpretation lies in the fact that when the prototype  $\xi^\mu$  is acquired, the synaptic weight  $w_{ij}$  should increase if neurons "i" and "j" receive a similar sign: in other words if  $\xi_i^\mu \cdot \xi_j^\mu$  is positive. On the other hand  $w_{ij}$  should decrease if  $\xi_i^\mu \cdot \xi_j^\mu$  is negative. The updating of the weights may be then expressed by,  $\Delta w_{ij} = \xi_i^\mu \cdot \xi_j^\mu$ , in other words, when the sign of the components "i" and "j" in the prototype  $\xi^\mu$  are with similar sign, the weight  $w_{ij}$  is positively reinforce, otherwise do the same but in a negative sense. In general a positive learning parameter  $\eta$  may be used, and it

can be state as the general training rule that the prototype  $\xi^\mu$  is stored then  $\Delta w_{ij} = \eta \cdot \xi_i^\mu \cdot \xi_j^\mu$  (being  $\eta$  a positive learning factor); which means that,  $\Delta w_{ij} \in \{-\eta, \eta\}$ . The synaptic matrix  $W$  should be interpreted as a sort of sign correlation matrix of the prototypes.

## 2. Training: Parameters of the Net

In our approach, instead of a matrix for storing the weights, a weight vector  $\bar{p} = (p_1, p_2, \dots, p_n)$  is going to be introduced. At the beginning  $\bar{p} = (0, 0, \dots, 0)$ . At time  $t$ , when the training pattern  $\xi^\mu$  is acquired, the weight vector  $\bar{p}$ , will be updated by this very simple rule:  $\bar{p} = \bar{p} + \xi^\mu$

which means that,  $\forall i, j = 1, \dots, n$  then,

$$\left\{ \begin{array}{l} \text{If } (\xi_i^\mu = 1 \text{ and } \xi_j^\mu = 1) \text{ then } (p_i = p_i + 1 \text{ and } p_j = p_j + 1) \\ \text{If } (\xi_i^\mu = -1 \text{ and } \xi_j^\mu = -1) \text{ then } (p_i = p_i - 1 \text{ and } p_j = p_j - 1) \\ \text{If } (\xi_i^\mu = -1 \text{ and } \xi_j^\mu = 1) \text{ then } (p_i = p_i - 1 \text{ and } p_j = p_j + 1) \\ \text{If } (\xi_i^\mu = 1 \text{ and } \xi_j^\mu = -1) \text{ then } (p_i = p_i + 1 \text{ and } p_j = p_j - 1) \end{array} \right.$$

It is clear that in this way training is faster than in the classical procedure, and the knowledge stored in the net parameter is equivalent. The synaptic matrix  $W = (w_{ij})$ , may be rebuilt by doing,

$$w_{ij} = p_i + p_j, \quad \forall i, j = 1, \dots, n$$

Which is equivalent to state that when the prototype  $\xi^\mu$  is acquired, the synaptic weight  $w_{ij}$  should increase if neurons "i" and "j" are in state 1, and will decrease if neurons "i" and "j" are in state -1. We may then consider that with this procedure, instead of storing some kind of sign correlation of the prototypes, like in the classical procedure was done, we are storing the correlation prototypes features. This method has also a very good property, and this is that, for any  $i, j, r, s \in \{1, \dots, n\}$ , then as,

$$p_i + p_j + p_r + p_s = p_i + p_s + p_r + p_j$$

one has that

$$w_{ij} + w_{rs} = w_{is} + w_{rj}$$

So, as the way the parameters are stores is related with the features of the pattern components, we are going to use in our approach the Boolean space  $\{0, 1\}^n$  instead of the bimodal space  $\{-1, 1\}^n$ . The training procedure will be then defines by

$$\Delta w_{ij} = \begin{cases} +1 & \text{if } \xi_i^\mu = \xi_j^\mu = 1, i \neq j, \\ -1 & \text{if } \xi_i^\mu = \xi_j^\mu = 0, i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

Two important advantages may be extracted from the above approach. The first advantage is that the space required for storing the parameters is lesser than in the classical one: the parameters may be stored in the weight vector, and then doing  $w_{ij} = p_i + p_j$ , to span it to the weight matrix if necessary. The second advantage is that the training may be much more easily understood using the next graphical interpretation of the training algorithm:



A  $n$  complete graph  $G$  may be introduced associated with the net. At the first step, a null value is assigned to all the edges  $a_{ij}$ , then, when a learning pattern  $\xi^\mu$  is acquired by the net, it is superposed over the graph  $G$ . The components,  $\{\xi_1^\mu, \dots, \xi_n^\mu\}$ , are going to be mapped over the vertices  $\{v_1, \dots, v_n\}$  of  $G$ . This mapping may be interpreted as a coloring of the edges in  $G$ , in such a way that, if  $\xi_i^\mu = \xi_j^\mu = 1$ , the edge  $a_{ij}$  (whose ending vertices are  $v_i$  and  $v_j$ ) will be colored with a certain color, for example red. On the other hand, if  $\xi_i^\mu = \xi_j^\mu = 0$ , then  $a_{ij}$  will be colored with a different color, as for example blue. The rest of the edges in  $G$  remain uncolored. Once this coloring has been done, the value assigned over the, also complete, graph of red edges are positively reinforced and the value assigned over the edges of the blue graph are negatively reinforced. The value over the rest of the edges remains unchanged. Once the pattern  $\xi^\mu$  is acquired, the colors are erased and we repeat the same color assignation with the next pattern to be acquired by the net, and so on. When every vector in the training pattern set has been integrated in the net, the training stage is finished, the resulting graph  $G$  has become edge-valued and its weight matrix is the synaptic matrix  $W = (w_{ij})$  of the net. Now if we define the basic matrix  $U^k = (u_{ij}^k)$ , where

$$\begin{cases} u_{ij} = 1 & \text{if } (i = k \text{ xor } j = k) \\ u_{ij} = 0 & \text{otherwise} \end{cases}$$

then, any synaptic matrix  $W$  is generated by the set of basic matrices  $\{U^1, U^2, \dots, U^n\}$ . In other words,

$$W = p_1 \cdot U^1 + \dots + p_n \cdot U^n.$$

### 3. Recall

For recalling a pattern from the net, the net should be colored with the color associated with that pattern, which may be interpreted as if the net had in a certain state  $x(t)$ . Then, it is clear that, we may define the energy pair number  $\{I(t), O(t)\}$ , where

$$I(t) = \frac{1}{2} (x(t) \cdot W \cdot x(t)^t)$$

represents the sum of the values of all the parameters  $w_{ij}$ , associated with all the edges colored in red, and if  $\bar{x}(t)$  is the symmetric vector of  $x(t)$ , then

$$O(t) = \frac{1}{2} (\bar{x}(t) \cdot W \cdot \bar{x}(t)^t)$$

represents the sum of the values of all the parameters  $w_{ij}$ , associated with all the edges colored in blue. Taking now into account that

$$W = p_1 \cdot U^1 + \dots + p_n \cdot U^n$$

one has that

$$\frac{1}{2} (x(t) \cdot W \cdot x(t)^t) = \frac{1}{2} x(t) (p_1 U^1 + \dots + p_n U^n) x(t)^t = p_1 \left[ \frac{1}{2} x(t) \cdot U^1 \cdot x(t)^t \right] + \dots + p_n \left[ \frac{1}{2} x(t) \cdot U^n \cdot x(t)^t \right]$$

and

$$\frac{I}{2}(\bar{x}(t) \cdot W \cdot \bar{x}(t)^t) = \frac{I}{2}\bar{x}(t)(p_1U^1 + \dots + p_nU^n)\bar{x}(t)^t = p_1\left[\frac{I}{2}\bar{x}(t) \cdot U^1 \cdot \bar{x}(t)^t\right] + \dots + p_n\left[\frac{I}{2}\bar{x}(t) \cdot U^n \cdot \bar{x}(t)^t\right]$$

In other words, if  $\forall i, j = 1, \dots, n$ , we do

$$\begin{cases} I_i = \frac{I}{2}x(t) \cdot U^i \cdot x(t)^t \\ O_i = \frac{I}{2}\bar{x}(t) \cdot U^i \cdot \bar{x}(t)^t \end{cases}$$

then

$$\begin{cases} I(t) = p_1 \cdot I_1(t) + \dots + p_n \cdot I_n(t) \\ O(t) = p_1 \cdot O_1(t) + \dots + p_n \cdot O_n(t) \end{cases}$$

but if  $n_1$  is the number of unit components of  $x(t)$  and if  $n_0$  is the number of the null ones (in other words  $n_1$  is the Hamming distance from  $x(t)$  to the zero vector), it is obvious that

$$I_i(t) = \begin{cases} n_1 - 1 & \text{if } x_i(t) = 1 \\ 0 & \text{if } x_i(t) = 0 \end{cases}, \quad \text{and} \quad O_i(t) = \begin{cases} n_0 - 1 & \text{if } x_i(t) = 0 \\ 0 & \text{if } x_i(t) = 1 \end{cases}$$

So, if  $i_1, i_2, \dots, i_{n_1}$ , and  $j_1, j_2, \dots, j_{n_0}$  are the places where the unit and null components of  $x(t)$ , are respectively located, the above equations could be written as

$$I(t) = (n_1 - 1)(p_{i_1} + p_{i_2} + \dots + p_{i_{n_1}}) \quad \text{and} \quad O(t) = (n_0 - 1)(p_{j_1} + p_{j_2} + \dots + p_{j_{n_0}})$$

which means that

$$\frac{I(t)}{n_1 - 1} + \frac{O(t)}{n_0 - 1} = k \quad (\text{where } k = p_1 + \dots + p_n).$$

In other words: all states  $x(t)$  sharing the same distance "j" (where  $j \in \{1, 2, \dots, n\}$ ), will verify the before equation. The states' space could be then classified in the  $n$  classes  $[1], [2], \dots, [j], \dots, [n]$ . If the pair  $\{I(t), O(t)\}$ , where defined as the energy pair (PE), associated to  $x(t)$ , one could state that all the states in the same class, has theirs PE's in the same energy line. On the other hand, we may define the relative weight of the neuron "i" when the net is in state  $x(t)$ , as the contribution of this neuron to the component  $I(t)$ , if  $x_i(t) = 1$ ; or as the contribution of this neuron to the component  $O(t)$ , if  $x_i(t) = 0$ . So, if  $x_i(t) = 1$ , we define the *relative weight*  $w_i(t)$  of the neuron "i" when the net is in state  $x(t)$  as:

$$w_i(t) = \frac{\sum_{j=1}^n x_j(t) \cdot w_{ij}}{I(t)}$$

and taking into account that,

$$\begin{aligned} \sum_{j=1}^n x_j(t) \cdot w_{ij} &= x_1(t)(p_i + p_1) + \dots + x_i(t) \cdot 0 + \dots + x_n(t)(p_i + p_n) = \\ &= (x_1(t) \cdot p_1 + \dots + x_n(t) \cdot p_n) + p_i(x_1(t) + \dots + x_n(t)) - 2 \cdot x_i(t) \cdot p_i = p \cdot x(t) + p_i(n_1 - 2) \end{aligned}$$

and

$$I(t) = (n_i - 1) \cdot p \cdot x(t)$$

$w_i(t)$  may be represented as

$$w_i(t) = \frac{I}{n_i - 1} + \frac{n_i - 2}{n_i - 1} \cdot \frac{p_i}{p \cdot x(t)}$$

and without difficult it could be also proved that if  $x_i(t) = 0$ , then

$$w_i(t) = \frac{I}{n_0 - 1} + \frac{n_0 - 2}{n_0 - 1} \cdot \frac{p_i}{p \cdot x(t)}$$

#### 4. Dynamic Equation

If in time  $t$  the state vector  $x(t)$  is in class  $[j]$ , then for any  $i$  from 1 to  $n$ , the dynamic equation is defined as

$$x_i(t+1) = f_h \left[ (f_b(x_i(t))) \cdot (w_i(t) - \theta_j) \right]$$

where  $f_h$  is the Heaviside step function and  $f_b$  is the function defined as  $f_b(x) = 2 \cdot x - 1$ , which achieves the transformation from the domain  $\{0, 1\}$  to the domain  $\{-1, 1\}$ . In other words: if  $x(t)$  is in class  $[j]$  and  $x_i(t) = 1$  the above expression could be written as

$$x_i(t+1) = f_h(w_i(t) - \theta_j)$$

which states that if  $w_i(t) < \theta_j$ , then  $x_i(t)$  changes its state from state  $x_i(t) = 1$  to  $x_i(t) = 0$ ; and otherwise  $x_i(t)$  doesn't change its state. On the other hand, if  $x(t)$  is in class  $[j]$  and  $x_i(t) = 0$ , the above expression could be written as

$$x_i(t+1) = f_h(\theta_j - w_i(t))$$

which states that if  $w_i(t) < \theta_j$ , then  $x_i(t)$  changes its state from state  $x_i(t) = 0$  to  $x_i(t) = 1$ ; and otherwise  $x_i(t)$  doesn't change its state. The value of  $\theta_j$  is the  $j$ -th component of a  $n$ -dimensional vector  $\bar{\theta}$  and is related with the class  $[j]$  to which the state  $x(t)$  belongs, and must not to be interpreted as a threshold for the " $i$ " unit (which will be assumed to be zero, whenever this hypothesis is not critical for the results we will to establish). It is clear that the lower we set the value of  $\theta_j$ , the more states in class  $[j]$  will have their relative weights greater than  $\theta_j$  which means that more fixed points the class  $[j]$  will have. The desirable values for the, so to be called, capacity vector parameter  $\bar{\theta} = (\theta_1, \dots, \theta_n)$ , may be obtained in an adaptive way. It can also be stated that the sum of the relative weights  $w_i(t)$  for the unit components of  $x(t)$  is equal to 2. The same could be proved for the null components. We have then that the relative weight vector  $w(t) = (w_1(t), w_2(t), \dots, w_n(t))$  associated to any state vector  $x(t)$  may also be interpreted as a sort of frequency distribution of probabilities. The reason is that

$$\sum_{i=1}^n w_i(t) = 4 \Rightarrow \sum_{i=1}^n \frac{1}{4} \cdot w_i(t) = 1$$

For any relative weight vector  $w(t)$ . The "uniform distribution vector" would be the one with all its components equal to  $4/n$ , which mean that the relative weight vector may be interpreted as a sort of frequency distribution of probabilities, this distribution may be considered as the relative weight vector associated to that state.

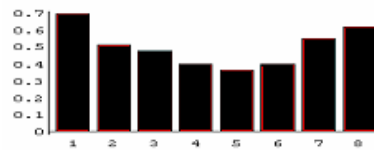


Figure 1 Relative Weight Vector associated to a certain state  $x$ , in a space of dimension 8.

### 5. Application

Our algorithm has been used in several applications. In this paper, we take, as an example for validating the performance of the algorithm we propose, the problem of the recognition of the Arabian digits as the prototype vectors:

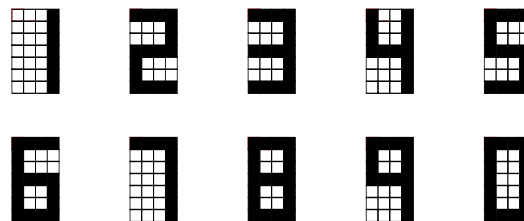


Figure 2 Arabian digits

Where the dimension  $n$ , of the pattern space is 28, and  $\xi^1 = [0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1]$ ,  $\xi^2 = [1,1,1,1,0,0,0,1,0,0,0,1,1,1,1,1,0,0,0,1,0,0,0,1,1,1,1]$ , and so on. After training, the weight vector is  $p = 1/14 \{53, 25, 25, 53, 11, -73, -73, 39, 11, -73, -73, 39, 39, 25, 25, 67, -17, -73, -73, 53, -17, -73, -73, 53, 11, 11, 11, 67\}$ . In figure 3 the reader may see the energy lines and theirs associated PE's. The Arabian digits are in this way placed on the lines:  $r_7, r_{16}, r_{16}, r_{13}, r_{16}, r_{15}, r_{10}, r_{20}, r_{15}, r_{18}$ . And the associated PE's are  $1/7\{1113,-3710\}, 1/7\{3420,-2508\}, 1/7\{4470,-3278\}, 1/7\{3210,-3745\}, 1/7\{4050,-2970\}, 1/7\{2821,-2418\}, 1/7\{2133,-4029\}, 1/7\{5548,-2044\}, 1/7\{4095,-3510\}, 1/7\{4539,-2403\}$ . The problem now is how to obtain in an adaptive way the capacity parameters  $\theta_1, \theta_2, \dots, \theta_{28}$ , in order to obtain the Arabian digits as fixed points with the least number of parasitic points as possible. When the before dynamic equation is considered, a point  $x(t)$  whose energy projection belongs to the  $r_j$  line, is a fixed point if, and only if, the (capacity) parameter  $\theta_j$  is an upper bound for all the relative weights  $w_i(t)$  associated to the components of  $x(t)$ .

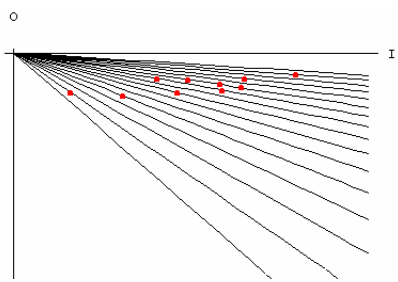


Figure 3 Arabian Digits Projections

Once the training has finished, the relative weight vector of the prototypes could then be calculated using. If the energy projection of the prototype  $\xi^\mu$  belongs to  $r_j$  and the largest of the components of  $w_i(t)$  is taken as  $\theta_j$ : it is clear that the prototype  $\xi^\mu$  will be a fixed point. But the problem is how to avoid that a point with high degree of correlation with a prototype but with all its relative weights components lower than the capacity parameter to skip away from this prototype. In our example, the number 2 is a prototype with 16 units components, in other words its energy projection is placed in the line  $r_{16}$  and the capacity parameter  $\theta_{16}$  is equal to 0.0741306.

If a little noise is added to the pattern (pattern in figure 4) and this noisily pattern (belonging to class  $r_{15}$ ) is given for retrieving:



Figure 4 Noisily Prototype 2 belonging to  $r_{15}$

It may happen that the noisily pattern changes to other state quite different from its natural attractor (the number 2) The reason for that, is that the prototype number 6, see figure 5, belongs also to the class  $r_{15}$  and the stability condition in this class was set very high.



Figure 5 Prototype 6 belonging to  $r_{15}$

The question is how to get that only the second component changes its state, when the Noisily Prototype 2, is given for retrieving. In other words, how to get all the neighbors (inside a give radius) of a prototype to be attracted by this prototype, see figure 6. The idea, proposed in this paper, made use of the deviation defined in [Giménez 2000]. When, in time  $t$ , the dynamic equation is applied to a component of the vector  $x(t)$ , this component will change its state not only if the relative weight  $w_i(t)$  is lower that the capacity parameter of its class. The deviation of the new state, in the case of change of state, must be similar to the deviation of the prototypes in the new class. The degree of similarity may be measured by a coefficient  $\mu$ . The coefficient  $\mu$  is handled in a dynamical way (the more is the time the higher is the coefficient).

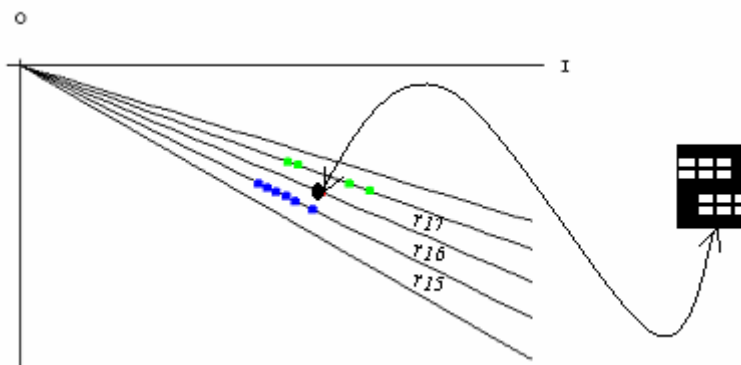


Figure 6 Neighbor of prototype 2 in  $r_{15}$  and  $r_{17}$

Besides the weight vector, there is other set of parameters of the net. For every one class  $r_i$ , the capacity parameter  $\theta_i$  and the deviation of the prototypes in this class are obtained. So the algorithm control not only if the new state is strongly correlate with some prototype in its class, the algorithm also control that the components in the new state must, with a high degree of probability, be placed in similar places as some prototype of the class. We have applied with to our example, obtaining that almost all the points inside a neighborhood of radius 1, of the prototypes, are attracted by these prototypes. The 10 Arabian digits are fixed points of the system, and almost all the 28 neighbor of any one of them were attracted by its attractor prototype. In figure 7, the number of points inside a neighborhood of radius 1, of the prototypes are expressed.

---

24	→	1	22	→	6
23	→	2	25	→	7
25	→	3	25	→	8
22	→	4	22	→	9
27	→	5	21	→	0

Figure 7 Prototype 6 belonging also to  $r_{15}$ 


---

## 6. Conclusion

The weight parameters in the Hopfield network are not a free set of variables. They must fulfill a set of constraints which have been deduced through a new re-interpretation of the net as Graph Formalisms. Making use of this constraint the state-vector has been classified in  $n$  classes according to the  $n$  different possible distances from any of the state-vectors to the zero vector. The  $(n \times n)$  matrix of weights may also be reduced to an  $n$ -vector of weights. In this way the computational time and the memory space, required for obtaining the weights, is optimized and simplified. The degree of correlation from a pattern with the prototypes may be controlled by the dynamical value of two parameters: the capacity parameter  $\theta$  which is used for controlling the capacity of the net (it may be proved that the bigger is the  $\theta_j$  component of  $\theta$ , the lower is the number of fixed points located in the  $r_j$  energy line) and the parameter  $\mu$  which measures the deviation to the prototypes. A typical example has been exposed; the obtained results have proved to improve the obtained when the classical algorithm is applied.

---

## Bibliography

- [Hopfield 82.] J. J. Hopfield. Neural Networks and physical systems with emergent collective behavior. Proc. Natl. Acad. Sci. USA, 79:2554, 1982.
- [Kinzel 85] W. Kinzel, Z. Phys. (B-Condensed Matter) Learning and pattern recognition in spin glass models, vol.60, pp.205-213, 1985
- [Elice 87] R. Mc. Elice, E. Posner, E. Rodemich, and S. Venkatesh. The capacity of the Hopfield associative memory. IEEE Trans. On Information Theory, vol.IT-33. pp.461-482, 1987.
- [Bose 96] N. K. Bose and P. Liang. Neural Network Fundamentals with Graphs, algorithms and Applications. McGraw Series in Electrical and Computer Engineering.1996.
- [Giménez 97] V. Giménez-Martínez, M. Pérez-Castellanos, J. Ríos Carrión and F. de Mingo, Capacity and Parasitic Fixed points Control in a Recursive Neural Network, Lecture Notes in Computer Science, SPRINGER-VERLAG, 1997, pp.215-226
- [Giménez 2000] V. Giménez-Martínez. A Modified Hopfield Algorithm Auto-Associative memory with Improved Capacity, IEEE Transactions on Neural Networks, vol.11,n.4, 2000, pp. 867-878.
- [Giménez 2001] Giménez-Martínez V., Erviti Anaut J. and Pérez-Castellanos M.M, Recurrent Neural Networks for Statistical Pattern Recognition, Frontiers in Artificial Intelligence and Applications, Vol.69, part 1, n.3, 2001, pp.1152-1159.
- [Giménez 2001] Giménez-Martínez V., Aslanyan L., Castellanos J.and Ryazanov V., Distribution functions as attractors for Recurrent Neural Networks Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, vol.11, n.3, 2001, pp.492-497.

---

## Authors' Information

**V. Giménez-Martínez** – Dep. de Matemática Aplicada.

**C. Torres** – Dep. de Matemática Aplicada.

**J. Erviti Anaut** – Dep. de Matemática Aplicada.

**M. Perez-Castellanos** – Dep de Arquitectura y Tecnología. de Sistemas

Facultad de Informática, U.P.M. Campus de Montegancedo s/n Boadilla del Monte, 28660 MADRID, SPAIN

e-mail: [vjimenez@fi.upm.es](mailto:vjimenez@fi.upm.es)

## ON CONTRADICTION DEGREES BETWEEN TWO FUZZY SETS, REVISITED

Carmen Torres, Elena Castiñeira, Susana Cubillo, Victoria Zarzosa

**Abstract:** Several methods have been proposed within fuzzy logic for inferring new knowledge from the original premises. However, there must be some guarantee that the results do not contradict each other nor they contradict the initial information. In [5] and [6], Trillas *et al.* introduced the concepts of both self-contradictory fuzzy set and contradiction between two fuzzy sets. Moreover, the need to study not only contradiction but also the degree of such contradiction is pointed out in [1] and [2] suggesting some measures for this purpose. Nevertheless, contradiction could have been measured in some other way. In [3] an axiomatic definition of measure of contradiction is proposed, both, for a set and between two sets. This paper continues along the lines of a study started out in [1] and [2] defining new degrees of contradiction between two fuzzy sets both depending on a given negation and without depending on any negation.

**Keywords:** fuzzy sets, *t*-norm, *t*-conorm, fuzzy strong negations, contradiction, measures of contradiction.

### Introduction

One of the main problems tackled by fuzzy logic is how to deal with inferences that include imprecise information. So, several methods have been proposed within this field for inferring new knowledge from the original premises. In any inference process, however, we have to assure that the results yielded neither contradict each other nor the original information.

The concept of contradiction in fuzzy logic was introduced by Trillas *et al.* in [5] and [6]. These papers formalize the idea that a fuzzy set  $P$  associated with a vague predicate  $\mathbf{P}$  is contradictory if it violates the principle of non-contradiction in the following sense: the statement "If  $x$  is  $\mathbf{P}$ , then  $x$  is **not**  $\mathbf{P}$ " holds with some degree of truth. So, they established that the fuzzy set  $P$  is contradictory if " $\mu_P(x) \rightarrow \mu_{\neg P}(x)$  for all  $x$  in the universe of discourse", where  $\mu_P(x)$  represents the degree in which  $x$  verifies the predicate  $\mathbf{P}$ . Also they represented the implication " $\rightarrow$ " by means of the inequality  $\leq$  of the real numbers, and the degree in which  $x$  verifies **not**  $\mathbf{P}$  as  $\mu_{\neg P} = N \circ \mu_P$  where  $N$  is a function representing an involutive negation; then  $P$  (or  $\mu_P$ ) is contradictory regarding the negation  $N$  or  $N$ -contradictory if  $\mu_P \leq N \circ \mu_P$ . Contradiction between two fuzzy sets was also introduced in [5] and [6]. Two fuzzy sets  $P$  and  $Q$  associated with vague predicates  $\mathbf{P}$  and  $\mathbf{Q}$ , respectively, are contradictory if one of them "implies" the negation of the other, that is, " $\mu_P(x) \rightarrow \mu_{\neg Q}(x)$  for all  $x$  in the universe of discourse".

So, if  $\mu_P$  and  $\mu_Q$  are the functions that represent the degrees in which an element verifies the predicates  $\mathbf{P}$  and  $\mathbf{Q}$ , respectively, then the fuzzy sets  $P$  and  $Q$  (or equivalently,  $\mu_P$  and  $\mu_Q$ ) are  $N$ -contradictory if the condition  $\mu_P(x) \leq (N \circ \mu_Q)(x)$  holds for all  $x$ . The need to speak not only of contradiction but also of degrees of contradiction was later raised in [1] and [2], where a function was considered for the purpose of determining (or measuring) the contradiction degree of a fuzzy set. Also, in [2] the authors proposed a function that appears to be suited for measuring the degree of contradiction between two fuzzy sets. However, many functions could be constructed for these purposes, and it is useful to specify what conditions a function must meet to be used as a measure of contradiction. Specifically, some axioms are needed to be able to decide whether a function is suitable for measuring the degree of contradiction. These axioms were established in [3].

In this work, we retake the study of the contradiction between two fuzzy sets, focusing on the problem from a geometrical perspective that suggests new ways of defining measures of contradiction. Therefore, after a geometrical study to determine what we will name regions of contradiction and non-contradiction, we will then define some functions by analyzing some of its properties.

---

## Preliminaries

---

Firstly, we will introduce a series of definitions and properties for their subsequent development in this article.

**Definition 2.1 ([7])** A fuzzy set (FS)  $P$ , in the universe  $X \neq \emptyset$ , is a set given as  $P = \{(x, \mu(x)) : x \in X\}$  such that, for all  $x \in X$ ,  $\mu(x) \in [0, 1]$ , and where the function  $\mu: X \rightarrow [0, 1]$  is called membership function. We denote  $\mathcal{F}(X)$  the set of all fuzzy sets on  $X$ .

**Definition 2.2**  $P \in \mathcal{F}(X)$  with membership function  $\mu \in [0, 1]^X$  is to be said a normal fuzzy set if  $\text{Sup}\{\mu(x) : x \in X\} = 1$ .

**Definition 2.3** A fuzzy negation (FN) is a non-increasing function  $N: [0, 1] \rightarrow [0, 1]$  with  $N(0) = 1$  and  $N(1) = 0$ . Moreover,  $N$  is a strong fuzzy negation if the equality  $N(N(y)) = y$  holds for all  $y \in [0, 1]$ .

The strong negations were characterized by Trillas in [4]. He showed that  $N$  is a strong negation if and only if, there is an order automorphism  $g$  in the unit interval (that is,  $g: [0, 1] \rightarrow [0, 1]$  is an increasing continuous function with  $g(0) = 0$  and  $g(1) = 1$ ) such that  $N(y) = g^{-1}(1 - g(y))$ , for all  $y \in [0, 1]$ ; from now on, let us denote  $N_g = g^{-1}(1 - g)$ . Furthermore, the only fixed point of  $N_g$  is  $n_g = g^{-1}(1/2)$ .

---

## Measuring $N_g$ -contradiction between Two Fuzzy Sets

---

For inference purposes in both classical and fuzzy logic, neither the information itself should be contradictory, nor should any of the items of available information contradict each other. The same applies to the given information and information ascertained by means of any chosen method of inference. This is the reason why the contradiction between two fuzzy sets was addressed. As mentioned above,  $\mu$  and  $\sigma$  are said to be  $N_g$ -contradictory if  $\mu(x) \leq N_g(\sigma(x))$  for all elements  $x$  in the universe of discourse, which is equivalent to  $\mu(x) \leq g^{-1}(1 - g(\sigma(x)))$  for all  $x$ , and also to  $\text{Sup}\{g(\mu(x)) + g(\sigma(x)) / x \in X\} \leq 1$ . Here again ascertaining whether two sets are contradictory will fall short of the mark, and a distinction should be made between any differing degrees of contradiction occurring in such situations. This problem was addressed for the first time in [1] and [2].

In this section, in order to study the degree of  $N_g$ -contradiction between two fuzzy sets  $P$  and  $Q$  (with membership functions  $\mu, \sigma \in [0, 1]^X$ , respectively) we consider the set  $\{(\mu(x), \sigma(x)) : x \in X\}$  as a subset of  $[0, 1]^2$  (we denote it by  $X_{\mu\sigma}$  to be short) and we will firstly analyze in what regions of  $[0, 1]^2$   $X_{\mu\sigma}$  must remain provided that  $\mu$  and  $\sigma$  are  $N_g$ -contradictory. The aim of this analysis is to find some relation suggesting the way of measuring the  $N_g$ -contradiction between two fuzzy sets. Secondly, we propose, some possible functions in order to measure the degrees of contradiction, bearing in mind the mentioned analysis.

### Regions of $N_g$ -contradiction

As mentioned above, given  $\mu, \sigma \in [0, 1]^X$  and a strong negation  $N_g$ , then  $\mu$  and  $\sigma$  are  $N_g$ -contradictory if and only if

$$\mu(x) \leq N_g(\sigma(x)) \quad \forall x \in X \Leftrightarrow \sigma(x) \leq N_g(\mu(x)) \quad \forall x \in X \Leftrightarrow g(\mu(x)) + g(\sigma(x)) \leq 1 \quad \forall x \in X$$

Above inequalities determine a curve in the unit square, with equation  $y_1 = N_g(y_2)$  or  $y_2 = N_g(y_1)$  or  $g(y_1) + g(y_2) = 1$ , that is the border between two regions: the contradictory sets remain in one of them, and the other one is a region free of contradiction. Let us see these regions in several particular cases and after that, the general case will be discussed.

#### (a) $N_s$ -contradiction with standard negation $N_s(y) = 1 - y$

Let  $N_s = 1 - \text{id}$  be the standard negation that is generated by  $g = \text{id}$ . Then  $\mu$  and  $\sigma$  are  $N_s$ -contradictory if and only if  $\mu(x) + \sigma(x) \leq 1$  for all  $x \in X$ , that is equivalent to  $X_{\mu\sigma} \subset \{(y_1, y_2) \in [0, 1]^2 : y_1 + y_2 \leq 1\}$  (see figure 1(a)).

#### (b) $N_g$ -contradiction with $g(y) = y^2$

The order automorphism  $g(y) = y^2$  determines the strong negation  $N_g(y) = \sqrt{1 - y^2}$ , and the sets  $\mu, \sigma \in [0, 1]^X$  are  $N_g$ -contradictory if and only if  $\mu(x) \leq \sqrt{1 - \sigma(x)^2}$ , that is equivalent to  $\mu(x)^2 + \sigma(x)^2 \leq 1$ . Therefore,  $\mu$  and  $\sigma$  are  $N_g$ -contradictory if and only if



$$X_{\mu\sigma} = \{(\mu(x), \sigma(x)) : x \in X\} \subset \{(y_1, y_2) \in [0,1]^2 : y_1^2 + y_2^2 \leq 1\}$$

Then,  $X_{\mu\sigma}$  must remain inside or on the circumference with center (0,0) and radius 1 (see figure 1 (b)).

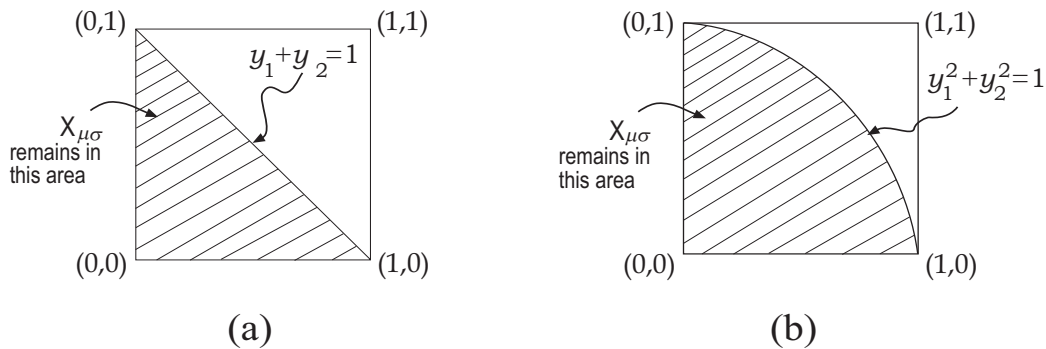


Figure 1:  $N_s$ -contradiction area and  $N_g$ -contradiction area with  $g(y)=y^2$ .

**(c) N-contradiction with N determined by  $g(y)=y^r, r>0$**

Let's consider the family of strong negations  $\{N_r\}_{r>0}$ , where for each  $r>0$  the automorphism that determines  $N_r$  is  $g_r(y)=y^r$ . This family includes as particular cases the negations given in (a) and (b) and for each  $r>0$  is  $N_r(y)=(1-y^r)^{1/r}$  with a fixed point  $y_{N_r} = \frac{1}{2^{1/r}}$ .  $\mu, \sigma \in [0,1]^X$  are  $N_r$ -contradictory if and only if

$$X_{\mu\sigma} \subset \{(y_1, y_2) \in [0,1]^2 : y_1^r + y_2^r \leq 1\}$$

For each  $r>0$  the curve  $y_1^r + y_2^r = 1$  is the border that delimits the region of contradiction, and if  $X_{\mu\sigma}$  takes some value  $(\mu(x_0), \sigma(x_0))$  over the mentioned curve, then, they are not  $N_r$ -contradictory.

We must note that as  $r$  increases, curves  $y_1^r + y_2^r = 1$  approach to the line  $y_1=1$  from the right (except in  $y_2=1$ ) and to the line  $y_2=1$  from above (except in  $y_1=1$ ); more specifically, the family of functions  $\{(1 - y_1^r)^{1/r}\}_{r>0}$  converges punctually when  $r \rightarrow \infty$ , to the constant function 1 for all  $y_1 \in [0,1)$  and in  $y_1=1$  converges to 0 and the family of functions  $\{(1 - y_2^r)^{1/r}\}_{r>0}$  converges punctually when  $r \rightarrow \infty$ , to the constant function 1 for all  $y_2 \in [0,1)$  and in  $y_2=1$  converges to 0; therefore, the region of non  $N_r$ -contradiction between two FS decreases when  $r$  grows (see figure 2). Moreover, when  $r \rightarrow 0$ , the family of functions  $\{(1 - y_2^r)^{1/r}\}_{r>0}$  converges for each  $y_2 \in (0,1]$  to the null function and for  $y_2=0$  converges to 1; and the family of functions  $\{(1 - y_1^r)^{1/r}\}_{r>0}$  converges for each  $y_1 \in (0,1]$  to the null function and for  $y_1=0$  converges to 1. That is, as  $r$  decreases the curves that delimit the regions of contradiction get closer to the axes  $y_1$  and  $y_2$ , and therefore, the region of non  $N_r$ -contradiction between two FS increases.

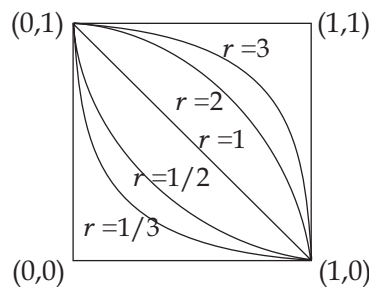


Figure 2: Curves  $y_1^r + y_2^r = 1$ .

On the other hand, if  $0 < r < s$  then, the curve  $y_1^s + y_2^s = 1$  is over curve  $y_1^r + y_2^r = 1$  (see figure 2 for the representation of some of them) and so, if  $\mu, \sigma \in [0,1]^X$  are  $N_r$ -contradictory, then they are  $N_s$ -contradictory for all  $s > r$ . In fact, if  $r < s$  it is  $y_1^r > y_1^s$  for all  $y_1 \in (0,1)$ , and therefore taking into account that  $g_{\frac{1}{r}}$  is increasing and that

$1/s < 1/r$ , is  $(1 - y_1^r)^{1/r} < (1 - y_1^s)^{1/r} < (1 - y_1^s)^{1/s}$  from where we follow that coordinate  $y_2$  of the curve corresponding to  $s$  is bigger than the one corresponding to  $r$ .

Finally, we observe that the family of curves mentioned above, practically fills the unit square  $[0,1]^2$  (with the exception of the border of the unit square except point  $(0,1)$  and  $(1,0)$ ), That is:

$$\bigcup_{r>0} \{(y_1, y_2) \in [0,1]^2 : y_1^r + y_2^r = 1\} = (0,1)^2 \cup \{(1,0), (0,1)\}$$

#### (d) General case of N-contradiction

If  $N$  is a strong FN, two sets  $\mu, \sigma \in [0,1]^X$  are  $N$ -contradictory if and only if

$$X_{\mu\sigma} \subset \{(y_1, y_2) \in [0,1]^2 : y_1 \leq N(y_2)\}$$

And the border curve that delimits the region exempt of contradiction is the curve of equation  $y_1 = N(y_2)$ . Therefore, the border curve is determined by a strong negation and it will have the following properties of a strong negation:

- 1) It is decreasing in both variable  $y_1$  and  $y_2$ .
- 2) It goes through  $(1,0)$  and through  $(0,1)$  since  $N(0)=1$  y  $N(1)=0$ .
- 3) It is symmetric with respect to the line  $y_1=y_2$  since  $y_1=N(y_2)$  and  $y_2=N(y_1)$  are the same curve, because  $N(N(y))=y$  for all  $y \in [0,1]$ .

Then, the regions of contradiction are limited by all strong negations in  $[0,1]$ .

#### Degrees of N-contradiction between Two Fuzzy Sets

As we discussed in the introduction, it is relevant to weight in which degree two sets are contradictory. In fact,  $\mu \emptyset \text{ y } \mu \emptyset$  (where  $\mu \emptyset(x)=0$  for all  $x \in X$ ) are  $N$ -contradictory for any strong FN  $N$ , and if we have two  $N$ -contradictory FS in a way, that at some point, one is the negation of the other, or what is the same,  $X_{\mu\sigma} \cap \{(y_1, y_2) \in [0,1]^2 : y_1 = N(y_2)\} \neq \emptyset$  (that is  $\mu(x_0)=N(\sigma(x_0))$  for some  $x_0 \in X$ ), small disturbances over value  $(\mu(x_0), \sigma(x_0))$  would result in very similar sets to the originals but not  $N$ -contradictory, meanwhile, small disturbances would never change the contradictory character of the empty set itself. Thus, it seems adequate that the degree of  $N$ -contradiction be 0 for whichever set  $\mu$  and  $\sigma$  such that  $X_{\mu\sigma} \cap \{(y_1, y_2) \in [0,1]^2 : y_1 \geq N(y_2)\} \neq \emptyset$ ; and that the degree be positive and it be as much higher as  $X_{\mu\sigma}$  is farther away from the limit curve  $(y_1=N(y_2))$ . Taking into account these observations, we are going to define different functions that could serve as a model to determine the different degrees of contradiction between two fuzzy sets

**Definition 3.1** Given  $\mu, \sigma \in [0,1]^X$ , we define the following contradiction measure functions:

- i)  $C_1^N(\mu, \sigma) = \text{Max}\left(0, \text{Inf}_{x \in X} (N(\sigma(x)) - \mu(x))\right)$
- ii)  $C_2^N(\mu, \sigma) = \text{Max}\left(0, \text{Inf}_{x \in X} (N(\mu(x)) - \sigma(x))\right)$
- iii)  $C_3^N(\mu, \sigma) = \text{Max}\left(0, 1 - \text{Sup}_{x \in X} (g(\mu(x)) + g(\sigma(x)))\right)$

- iv)  $C_4^N(\mu, \sigma) = 0$  if  $\mu$  and  $\sigma$  are not  $N$ -contradictory, and in the opposite case  $C_4^N(\mu, \sigma) = \frac{d(X_{\mu\sigma}, L_N)}{d((0,0), L_N)}$

where  $d$  is the Euclidean distance and  $L_N = \{(y_1, y_2) \in [0,1]^2 : N(y_1) = y_2\}$  is the limit curve, and

therefore,  $d(X_{\mu\sigma}, L_N) = \text{Inf} \{d((\mu(x), \sigma(x)), (y_1, y_2)) : x \in X, (y_1, y_2) \in L_N\}$  and  
 $d((0,0), L_N) = \text{Inf} \{d((0,0), (y_1, y_2)) : (y_1, y_2) \in L_N\}$ .

**Observation:** The four previous functions take values in  $[0,1]$  and it is verified that all of them are zero or all are strictly positive. The functions  $C_1^N$  and  $C_2^N$  come motivated by the characterization of contradiction " $\mu$  and  $\sigma$  are  $N_g$ -contradictory if and only if  $\mu(x) \leq N_g(\sigma(x)) \forall x \in X \Leftrightarrow \sigma(x) \leq N_g(\mu(x)) \forall x \in X$ ", while  $C_3^N$  is based on the characterization " $\mu$  and  $\sigma$  are  $N_g$ -contradictory if and only if  $g(\mu(x)) + g(\sigma(x)) \leq 1 \forall x \in X$ ". Although both characterizations are equivalent,  $C_1^N$ ,  $C_2^N$  y  $C_3^N$  they do not coincide, as the following examples will show. On the other hand,  $C_4^N$  represents a relative distance; the Euclidean distance of the set that the two FS describe to the limit curve relative to the distance of the "most contradictory" sets to the same curve. While  $C_1^N$  represents the infimum of the distances between the abscises of the values  $(\mu(x), \sigma(x))$  and the corresponding of the limit curve (see figure 3),  $C_2^N$  represents the infimum of the distances between the ordinates of the values  $(\mu(x), \sigma(x))$  and the corresponding of the limit curve (see figure 3). As far as  $C_3^N$  is concerned, some geometrical interpretations can be found in some particular cases.

**Proposition 3.2** Let  $N_{id}$  be the standard FN then for all  $\mu, \sigma \in [0,1]^X$  the degrees of contradiction between  $\mu$  and  $\sigma$  by means of the formula in definition 3.1 verify that  $C_1^{N_{id}}(\mu, \sigma) = C_2^{N_{id}}(\mu, \sigma) = C_3^{N_{id}}(\mu, \sigma) = C_4^{N_{id}}(\mu, \sigma)$

**Proposition 3.3** Let  $N_g$  be the strong FN with  $g(y)=y^2$ , i.e.  $N_g(y) = \sqrt{1-y^2}$ , then for all  $\mu, \sigma \in [0,1]^X$  the degrees of contradiction  $C_3^{N_g}$  and  $C_4^{N_g}$  between  $\mu$  and  $\sigma$  verify that  $C_3^{N_g}(\mu, \sigma) = 1 - (1 - C_4^{N_g}(\mu, \sigma))^2$ .

However, generally, the four measures are different as the following examples show.

**Example 3.4** Given  $P, Q \in \mathcal{F}(X)$  fuzzy sets over a finite universe, that is, their functions of membership  $\mu$  and  $\sigma$  take a finite number of values in  $[0,1]$ . Then,  $P$  and  $Q$  define a finite set of points in  $[0,1]^2$ , let it be this set  $X_{\mu\sigma} = \{(x_1, y_1), \dots, (x_n, y_n)\}$  being  $\{x_1, \dots, x_n\}$  and  $\{y_1, \dots, y_n\}$  the values that the functions of membership  $\mu$  and  $\sigma$  take respectively.

(i) Let's suppose  $\mu(X) = \{0.25, 0.5, 0.79\}$  and  $\sigma(X) = \{0.75, 0.63, 0.06\}$ . We consider the strong fuzzy negation  $N_g(y) = \sqrt{1-y^2}$ , with  $g(y)=y^2$ . Then:

$$C_1^{N_g}(\mu, \sigma) = \text{Max} \left( 0, \text{Inf}_{x \in X} \left( \sqrt{1 - \sigma(x)^2} - \mu(x) \right) \right) = 0.208 \text{ reaching the infimum at point } (0.79, 0.06)$$

$$C_2^{N_g}(\mu, \sigma) = \text{Max} \left( 0, \text{Inf}_{x \in X} \left( \sqrt{1 - \mu(x)^2} - \sigma(x) \right) \right) = 0.218 \text{ reaching the infimum at point } (0.25, 0.75)$$

$$C_3^{N_g}(\mu, \sigma) = \text{Max} \left( 0, 1 - \text{Sup}_{x \in X} \left( \mu(x)^2 + \sigma(x)^2 \right) \right) = 0.353 \text{ reaching the supremum at point } (0.5, 0.63)$$

Since  $\mu$  and  $\sigma$  are  $N_g$ -contradictory, then

$$C_4^{N_g}(\mu, \sigma) = \frac{d(X_{\mu\sigma}, L_N)}{d((0,0), L_N)} = d(X_{\mu\sigma}, \{(y_1, y_2) \in [0,1]^2 : y_1^2 + y_2^2 = 1\}) = 0.196$$

reaching the infimum of the distances at the same point as the previous case. Besides, as it can be seen, it is verified that:  $C_3^{N_g}(\mu, \sigma) = 0.353 = 1 - (1 - C_4^{N_g}(\mu, \sigma))^2 = 1 - (1 - 0.196)^2$

(ii) Let's suppose now  $\mu(X) = \{0.25, 0.63, 0.79, 0.79\}$  and  $\sigma(X) = \{0.75, 0.64, 0.06, 0.1\}$  and we consider the strong fuzzy negation  $N_g(y) = (1 - y^3)^{1/3}$ , with  $g(y)=y^3$ . Then:

$$C_1^{N_g}(\mu, \sigma) = \text{Inf}_{x \in X} \left( (1 - \sigma(x)^3)^{1/3} - \mu(x) \right) = 0.2096 \text{ reaching the infimum at point } (0.79, 0.1)$$

$$C_2^{N_g}(\mu, \sigma) = \inf_{x \in X} \left( (1 - \mu(x)^3)^{1/3} - \sigma(x) \right) = 0.2447 \text{ reaching the infimum at point } (0.25, 0.75)$$

$$C_3^{N_g}(\mu, \sigma) = 1 - \sup_{x \in X} (\mu(x)^3 + \sigma(x)^3) = 0.4878 \text{ reaching the supremum at point } (0.63, 0.64)$$

Since  $\mu$  and  $\sigma$  are  $N_g$ -contradictory and since  $L_N = \{(y_1, y_2) \in [0,1]^2 : y_1^3 + y_2^3 = 1\}$ , then

$$C_4^{N_g}(\mu, \sigma) = \frac{d(X_{\mu\sigma}, L_N)}{d((0,0), L_N)} = d((0.79, 0.06), L_N) = 0.218$$

In this case, it can be showed that

$$C_3^{N_g}(\mu, \sigma) = 0.4878 \neq 1 - (1 - C_4^{N_g}(\mu, \sigma))^2 = 1 - (1 - 0.218)^2 = 0.3884$$

**Example 3.5** Given  $P, Q \in \mathcal{F}[0,1]$  with membership functions  $\mu(x) = \frac{3x}{4}$  and  $\sigma(x) = \frac{x}{2}$ , we consider the strong fuzzy negation  $N_g(y) = \sqrt{1 - y^2}$ , with  $g(y) = y^2$ . Then:

$$C_1^{N_g}(\mu, \sigma) = \max \left( 0, \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{x}{2}\right)^2} - \frac{3x}{4} \right) \right) = 0.116 \text{ reaching the infimum at } x=1$$

$$C_2^{N_g}(\mu, \sigma) = \max \left( 0, \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{3x}{4}\right)^2} - \frac{x}{2} \right) \right) = 0.161 \text{ reaching the infimum at the same } x=1$$

$$C_3^{N_g}(\mu, \sigma) = \max \left( 0, 1 - \sup_{x \in [0,1]} \left( \left(\frac{3x}{4}\right)^2 + \left(\frac{x}{2}\right)^2 \right) \right) = \frac{3}{16} \text{ reaching the supremum at } x=1 \text{ as well.}$$

since  $\mu$  and  $\sigma$  are  $N_g$ -contradictory, then  $C_4^{N_g}(\mu, \sigma) = \frac{d(X_{\mu\sigma}, L_N)}{d((0,0), L_N)} = 1 - \frac{\sqrt{13}}{4}$ .

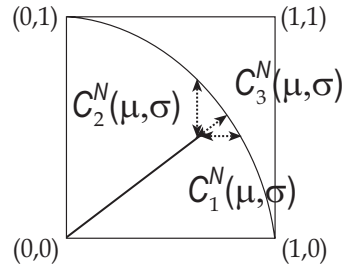


Figure 3: Geometrical interpretation of different contradiction degrees

Let's suppose now another set  $M \in \mathcal{F}[0,1]$  with membership function  $\eta(x) = \frac{3}{4} - \frac{3x}{4}$ , then:

$$C_1^{N_g}(\eta, \sigma) = \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{x}{2}\right)^2} - \left(\frac{3}{4} - \frac{3x}{4}\right) \right) = \frac{1}{4} \text{ reaching the infimum at } x=0,$$

$$C_2^{N_g}(\eta, \sigma) = \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{3}{4} - \frac{3x}{4}\right)^2} - \left(\frac{x}{2}\right) \right) = \frac{1}{2} \text{ reaching the infimum at } x=1,$$

$$C_3^{N_g}(\eta, \sigma) = 1 - \sup_{x \in [0,1]} \left( \left(\frac{3}{4} - \frac{3x}{4}\right)^2 + \left(\frac{x}{2}\right)^2 \right) = \frac{7}{16} \text{ reaching the supremum at } x=0.$$

Since  $\eta$  and  $\sigma$  are  $N_g$ -contradictory, then  $C_4^{N_g}(\eta, \sigma) = d(X_{\eta\sigma}, L_N) = 1 - \frac{3}{4}$ .

Lastly, let's suppose another set  $N \in \mathcal{A}[0,1]$  with membership function  $\delta(x) = \frac{x}{2} + \frac{1}{3}$ , then:

$$C_1^{N_g}(\eta, \delta) = \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{x}{2} + \frac{1}{3}\right)^2} - \left(\frac{3}{4} - \frac{3x}{4}\right) \right) = 0.192 \text{ reaching the infimum at } x=0,$$

$$C_2^{N_g}(\eta, \delta) = \inf_{x \in [0,1]} \left( \sqrt{1 - \left(\frac{3}{4} - \frac{3x}{4}\right)^2} - \left(\frac{x}{2} + \frac{1}{3}\right) \right) = \frac{1}{6} \text{ reaching the infimum at } x=1,$$

$$C_3^{N_g}(\eta, \delta) = 1 - \sup_{x \in [0,1]} \left( \left(\frac{3}{4} - \frac{3x}{4}\right)^2 + \left(\frac{x}{2} + \frac{1}{3}\right)^2 \right) = \frac{11}{36} \text{ reaching the supremum at } x=1.$$

Since  $\eta$  and  $\delta$  are  $N_g$ -contradictory, then  $C_4^{N_g}(\eta, \delta) = d(X_{\eta\delta}, L_N) = \frac{1}{6}$ .

The following properties of the above measure of N-contradiction functions between two fuzzy sets can be proved.

**Proposition 3.6** For each  $i=1,2,3,4$  function  $C_i^N : [0,1]^X \times [0,1]^X \rightarrow [0,1]$  defined for every two  $\mu, \sigma \in [0,1]^X$  as definition 3.1 verifies:

- i)  $C_i^N(\mu_\emptyset, \mu_\emptyset) = 1$ .
- ii)  $C_i^N(\mu, \sigma) = 0$  if  $\mu$  or  $\sigma$  normal.
- iii) Symmetry:  $C_i^N(\mu, \sigma) = C_i^N(\sigma, \mu)$  for  $i=3,4$ . For  $i=1,2$  is verified that  $C_1^N(\mu, \sigma) = C_2^N(\sigma, \mu)$ .
- iv) Given  $\{\mu_\alpha\}_{\alpha \in I} \subset [0,1]^X$ , it holds that:  $\inf_{\alpha \in I} C_i^N(\mu_\alpha, \sigma) = C_i^N\left(\sup_{\alpha \in I} \mu_\alpha, \sigma\right)$ . As a particular case of (iv) it is

verified that given  $\mu_1, \mu_2, \sigma \in [0,1]^X$  if  $\mu_1 \leq \mu_2$ , then  $C_i^N(\mu_1, \sigma) \geq C_i^N(\mu_2, \sigma)$  (Anti-Monotonicity).

Property (ii) is stronger than the second axiom given in [3] ( $C(\mu, \mu) = 0$  for all normal  $\mu \in [0,1]^X$ ) to define measures of contradiction. Moreover,  $C_i^N$  for each  $i=1,2,3,4$ , is a positive or strict measure of contradiction as defined in [3] since  $C_i^N(\mu, \sigma) = 0$  provided that  $\sup_{x \in X} (g(\mu(x)) + g(\sigma(x))) \geq 1$ .

**Example 3.7** Given  $P, Q \in \mathcal{A}[0,1]$  with membership functions  $\mu, \sigma$  such that  $\mu(x) = -2x+1$  if  $x \leq 1/2$  and 0, if  $x > 1/2$  and  $\sigma(x) = x$ , if  $x \leq 1/2$  and  $1/2$ , if  $x > 1/2$ . As  $\mu$  is normal, for all strong negation  $N$  the degree of contradiction is zero,  $C_i^N(\mu, \sigma) = 0$  with  $i=1,2,3,4$ . However, there are strong negations for which sets  $P, Q$  are contradictory. For instance, for all negations  $N_g$  such that  $g(y) = y^p$  with  $p \geq 1$ .

## Measuring Contradiction between Two Fuzzy Sets

In this section, we will deal with the case of contradiction without depending on a prefixed negation. The previous section establishes the contradiction between two fuzzy sets related to a chosen strong negation. We now address contradiction more generally, without depending on any specific FN. In [5] and [6] two FS  $P, Q \in \mathcal{A}(X)$  with membership functions  $\mu, \sigma$  were defined contradictory if they were N-contradictory regarding some strong FN  $N$ . The following result was proved in [2].

**Proposition 4.1 ([2])** If  $P, Q \in \mathcal{A}(X)$  with membership functions  $\mu, \sigma$  are contradictory, then: for all  $\{x_n\}_{n \in \mathbb{N}} \subset X$ , if  $\lim_{n \rightarrow \infty} \{\mu(x_n)\} = 1$ , then  $\lim_{n \rightarrow \infty} \{\sigma(x_n)\} = 0$ , and if  $\lim_{n \rightarrow \infty} \{\sigma(x_n)\} = 1$ , then  $\lim_{n \rightarrow \infty} \{\mu(x_n)\} = 0$ . In particular, if  $\mu(x) = 1$ , for some  $x \in X$ , then  $\sigma(x) = 0$  and vice-versa.

With the intention of measuring how contradictory two FS are, we will define some functions motivated in the previous section, being of interest, for one of them, to consider the following corollary also given in [2].

**Corollary 4.2 ([2])** If  $\mu, \sigma \in [0,1]^X$  are contradictory, then  $\text{Sup}_{x \in X}(\mu(x) + \sigma(x)) < 2$ .

**Definition 4.3** Given  $\mu, \sigma \in [0,1]^X$ , we define the following contradiction measure functions:

- i)  $C_1(\mu, \sigma) = 0$  if there exists  $\{x_n\}_{n \in \mathbb{N}} \subset X$  such that  $\lim_{n \rightarrow \infty} \{\mu(x_n)\} = 1$  or  $\lim_{n \rightarrow \infty} \{\sigma(x_n)\} = 1$ , and, in other case

$$C_1(\mu, \sigma) = \text{Min} \left( \text{Inf}_{x \in X} (1 - \mu(x)), \text{Inf}_{x \in X} (1 - \sigma(x)) \right).$$

- ii)  $C_2(\mu, \sigma) = 0$  if there exists  $\{x_n\}_{n \in \mathbb{N}} \subset X$  such that  $\lim_{n \rightarrow \infty} \{\mu(x_n)\} = 1$  or  $\lim_{n \rightarrow \infty} \{\sigma(x_n)\} = 1$ , and, in other case

$$C_2(\mu, \sigma) = 1 - \frac{\text{Sup}_{x \in X} (\mu(x) + \sigma(x))}{2}.$$

**Observation:** It is evident that the function  $C_1$  measures the minimum between distance (Euclidean) of  $X_{\mu\sigma}$  to the line  $y_1=1$  (that we will note  $L_1$ ) and the distance of  $X_{\mu\sigma}$  to the line  $y_2=1$  (that we will note  $L_2$ ):

$$C_1(\mu, \sigma) = \text{Min}(d(X_{\mu\sigma}, L_1), d(X_{\mu\sigma}, L_2)) = \frac{\text{Min}(d(X_{\mu\sigma}, L_1), d(X_{\mu\sigma}, L_2))}{d((0,0), L_1 \cap L_2)}.$$

On the other hand,  $C_2(\mu, \sigma) = \frac{d_1(X_{\mu\sigma}, (1,1))}{2} = \frac{d_1(X_{\mu\sigma}, (1,1))}{d_1((0,0), (1,1))}$  that is, the function  $C_2$  measures the reticular

distance between  $X_{\mu\sigma}$  and  $(1,1)$ , relative to the reticular distance from  $(0,0)$  to  $(1,1)$  (let us remind that  $d_1((y_1, y_2), (z_1, z_2)) = |y_1 - z_1| + |y_2 - z_2|$ ). These geometrical interpretations of the measures  $C_1$  and  $C_2$  suggest another way of measuring the contradiction degree:  $C_3(\mu, \sigma) = 0$  if there exists  $\{x_n\}_{n \in \mathbb{N}} \subset X$  such that

$$\lim_{n \rightarrow \infty} \{\mu(x_n)\} = 1 \text{ or } \lim_{n \rightarrow \infty} \{\sigma(x_n)\} = 1, \text{ and, in other case } C_3(\mu, \sigma) = \frac{d(X_{\mu\sigma}, (1,1))}{d((0,0), (1,1))}.$$

In the same way that happened with measures of N-contradiction between two fuzzy sets, the following result can be demonstrated.

**Proposition 4.4** For each  $i=1,2,3$  function  $C_i: [0,1]^X \times [0,1]^X \rightarrow [0,1]$  defined for each pair  $\mu, \sigma \in [0,1]^X$  as the above definition verifies:

- i)  $C_i(\mu_\emptyset, \mu_\emptyset) = 1$ .
- ii)  $C_i(\mu, \sigma) = 0$  if  $\mu$  or  $\sigma$  normal.
- iii) Symmetry:  $C_i(\mu, \sigma) = C_i(\sigma, \mu)$ .
- iv) Anti-Monotonicity: given  $\mu_1, \mu_2, \sigma \in [0,1]^X$  if  $\mu_1 \leq \mu_2$ , then  $C_i(\mu_1, \sigma) \geq C_i(\mu_2, \sigma)$ . Besides, for the case  $i=1$  axiom of the infimum given in [3] is also verified. That is, given  $\{\mu_\alpha\}_{\alpha \in I} \subset [0,1]^X$ , it holds that:

$$\text{Inf}_{\alpha \in I} C_i(\mu_\alpha, \sigma) = C_i \left( \text{Sup}_{\alpha \in I} \mu_\alpha, \sigma \right).$$

## Bibliography

- [1] E. Castiñeira, S. Cubillo and S. Bellido. Degrees of Contradiction in Fuzzy Sets Theory. Proceedings IPMU'02, 171-176. Annecy (France), 2002.
- [2] E. Castiñeira, S. Cubillo. and S. Bellido. Contradicción entre dos conjuntos. Actas ESTYLF'02, 379-383. León (Spain), 2002, (in Spanish).
- [3] S. Cubillo and E. Castiñeira. Measuring contradiction in fuzzy logic. International Journal of General Systems, Vol. 34, Nº1, 39-59, 2004.
- [4] E. Trillas. Sobre funciones de negación en la teoría de conjuntos difusos. Stochastica III/1, 47-60, 1979 (in Spanish). Reprinted (English version) (1998) in Avances of Fuzzy Logic. Eds. S. Barro et alr, 31-43.
- [5] E. Trillas, C. Alsina and J. Jacas. On Contradiction in Fuzzy Logic. Soft Computing, 3(4), 197-199, 1999.
- [6] E. Trillas and S. Cubillo. On Non-Contradictory Input/Output Couples in Zadeh's CRI. Proceedings NAFIPS, 28-32. New York, 1999.
- [7] L. A. Zadeh. Fuzzy Sets. Inf. Control, volume 20, pages 301-312, 1965.

---

### Authors' Information

---

**Carmen Torres** – Dept. Applied Mathematic. Computer Science School of University Politécnica of Madrid. Campus Montegancedo. 28660 Boadilla del Monte (Madrid). Spain; e-mail: [ctorres@fi.upm.es](mailto:ctorres@fi.upm.es)

**Elena Castiñeira** – Dept. Applied Mathematic. Computer Science School of University Politécnica of Madrid. Campus Montegancedo. 28660 Boadilla del Monte (Madrid). Spain; e-mail: [ecastineira@fi.upm.es](mailto:ecastineira@fi.upm.es)

**Susana Cubillo** – Dept. Applied Mathematic. Computer Science School of University Politécnica of Madrid. Campus Montegancedo. 28660 Boadilla del Monte (Madrid). Spain; e-mail: [scubillo@fi.upm.es](mailto:scubillo@fi.upm.es)

**Victoria Zarzosa** – Dept. Applied Mathematic. Computer Science School of University Politécnica of Madrid. Campus Montegancedo. 28660 Boadilla del Monte (Madrid). Spain; e-mail: [vzarzosa@fi.upm.es](mailto:vzarzosa@fi.upm.es)

## AN APPROACH TO COLLABORATIVE FILTERING BY ARTMAP NEURAL NETWORKS

Anatoli Nachev

**Abstract:** Recommender systems are now widely used in e-commerce applications to assist customers to find relevant products from the many that are frequently available. Collaborative filtering (CF) is a key component of many of these systems, in which recommendations are made to users based on the opinions of similar users in a system. This paper presents a model-based approach to CF by using supervised ARTMAP neural networks (NN). This approach deploys formation of reference vectors, which makes a CF recommendation system able to classify user profile patterns into classes of similar profiles. Empirical results reported show that the proposed approach performs better than similar CF systems based on unsupervised ART2 NN or neighbourhood-based algorithm.

**Keywords:** neural networks, ARTMAP, collaborative filtering

---

### Introduction

---

The World Wide Web has been established as a major platform for information and application delivery. The amount of content and functionality available often exceeds the cognitive capacity of users. This problem has also been characterized as information overload [13]. Since the World Wide Web has become widespread, more and more applications exist that are suitable for the application of social information filtering techniques. Recommender systems are now widely used in e-commerce applications to assist customers to find relevant products from the many that are frequently available. Collaborative filtering is a key component of many of these systems, in which recommendations are made to users based on the opinions of similar users in a system.

In collaborative filtering preferences of a user are estimated through mining data available about the whole user population, implicitly exploiting analogies between users that show similar characteristics.

A variety of CF filters or recommender systems have been designed, most of which can be grouped into two major classes: memory-based and model-based [10].

Memory-based algorithms maintain a database of all users' known preferences for all items, and for each prediction, perform some computation across the entire database. This approach is simpler, seem to work reasonably well in practice, and new data can be added easily and incrementally, however, it can become computationally expensive in terms of both time and space complexity, as the size of the database grows.

On the other hand, model-based CF algorithms use the users' preferences to learn a model, which is then used for predictions. They are small, fast, and essentially as accurate as memory based methods. Memory requirements for the model are generally less than for storing the full database and predictions can be calculated quickly once the model is generated.

This paper presents a model based-approach to collaborative filtering by using supervised ARTMAP neural network. Proposed algorithm is based on formation of reference vectors that make a CF system able to classify user profile patterns into classes of similar profiles, which forms the basis of a recommendation system.

### Related Work

A variety of collaborative filters or recommender systems have been designed and deployed. The Tapestry system relied on each user to identify like-minded users manually [5]. GroupLens [6] and Ringo [7], developed independently, were the first CF algorithms to automate prediction. Both are examples of the more general class of memory-based approaches, where for each prediction, some measure is calculated over the entire database of users' ratings. Typically, a similarity score between the active user and every other user is calculated. Predictions are generated by weighting each user's ratings proportionally to his or her similarity to the active user. A variety of similarity metrics is possible. Resnick et al. [6] employ the Pearson correlation coefficient. Shardanand and Maes [7] test a few metrics, including correlation and mean squared difference. Breese et al. [8] propose the use of vector similarity, based on the vector cosine measure often employed in information retrieval. All of the memory-based algorithms cited predict the active user's rating as a similarity-weighted sum of the others users' ratings, though other combination methods, such as a weighted product, are equally plausible. Basu et al. [9] explore the use of additional sources of information (for example, the age or sex of users, or the genre of movies) to aid prediction. Breese et al. [8] identify a second general class of model-based algorithms. In this approach, an underlying model of user preferences is first constructed, from which predictions are inferred. The authors describe and evaluate two probabilistic models, which they term the Bayesian clustering and Bayesian network models.

### Adaptive Resonance Theory

Adaptive Resonance Theory (ART) [1] [2] is family of neural networks for fast learning, pattern recognition, and prediction, including both unsupervised: ART1, ART2, ART2-A, ART3, Fuzzy ART, Distributed ART; and supervised: ARTMAP, Fuzzy ARTMAP, ART-EMAP, ARTMAP-IC, ARTMAP-FTR, Distributed ARTMAP, and Default ARTMAP systems.

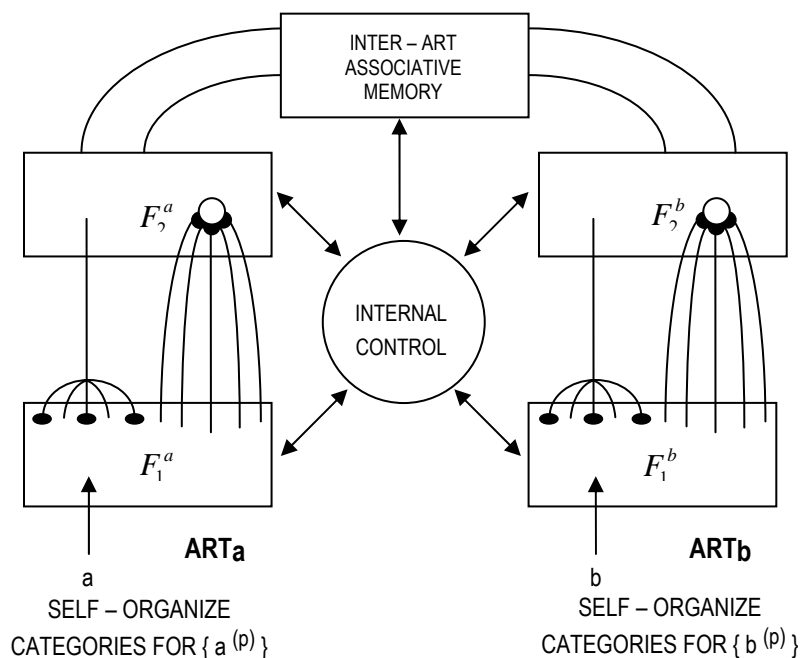


Figure 1. Components of an ARTMAP system.



These ART models have been used for a wide range of applications, such as remote sensing, medical diagnosis, automatic target recognition, mobile robots, and database management. ART1 self-organizes recognition codes for binary input patterns; ART2 does the same for analogue input patterns. ART3 is the same as ART2 but includes a model of the chemical synapse that solves the memory-search problem of ART systems. Any ART module consists of two fields,  $F_1$  and  $F_2$ , connected by two sets of adaptive connections: bottom-up connections,  $F_1 \rightarrow F_2$ ; and top-down connections  $F_2 \rightarrow F_1$ . In an ART module, the input pattern is presented to the  $F_1$  field which normalizes and contrast-enhances features of the pattern.  $F_2$  activation is then calculated by multiplying the  $F_1$  pattern with the bottom-up weights. Lateral inhibition in the  $F_2$  field then finds a winning  $F_2$  node. The degree of match between the top-down expectation pattern of the winning  $F_2$  node and the  $F_1$  pattern is then evaluated in a vigilance test to determine whether it is sufficient. If it is, then learning occurs in both the top-down and bottom-up connections of the winning  $F_2$  node, otherwise the winning  $F_2$  node is reset and the search continues. ARTMAP is a supervised neural network which consists of two unsupervised ART modules, ART<sub>a</sub> and ART<sub>b</sub> and an inter-ART associative memory, called a map-field (see Figure 1).

### ARTMAP Network

ARTMAP architectures are neural networks that develop stable recognition codes in real time in response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system has to face: how to keep learning from new events without forgetting previously learned information. ARTMAP networks were designed to accept binary input patterns [3].

An ART module has three layers: the input layer ( $F_0$ ), the comparison layer ( $F_1$ ), and the recognition layer ( $F_2$ ) with  $m$ ,  $m$  and  $n$  neurons, respectively (see module ART<sub>a</sub> or ART<sub>b</sub> in Figure 2). The neurons, or nodes, in the  $F_2$  layer represent input categories. The  $F_1$  and  $F_2$  layers interact with each other through weighted bottom-up and top-down connections, which are modified when the network learns. There are additional gain control signals in the network that regulate its operation.

At each presentation of a non-zero binary input pattern  $x$  ( $x \in \{0,1\}, i=1,2,\dots,m$ ), the network attempts to classify it into one of its existing categories based on its similarity to the stored prototype of each category node. More precisely, for each node  $j$  in the  $F_2$  layer, the bottom-up activation

$$T_j = \sum_{i=1}^m x_i Z_{ij}$$

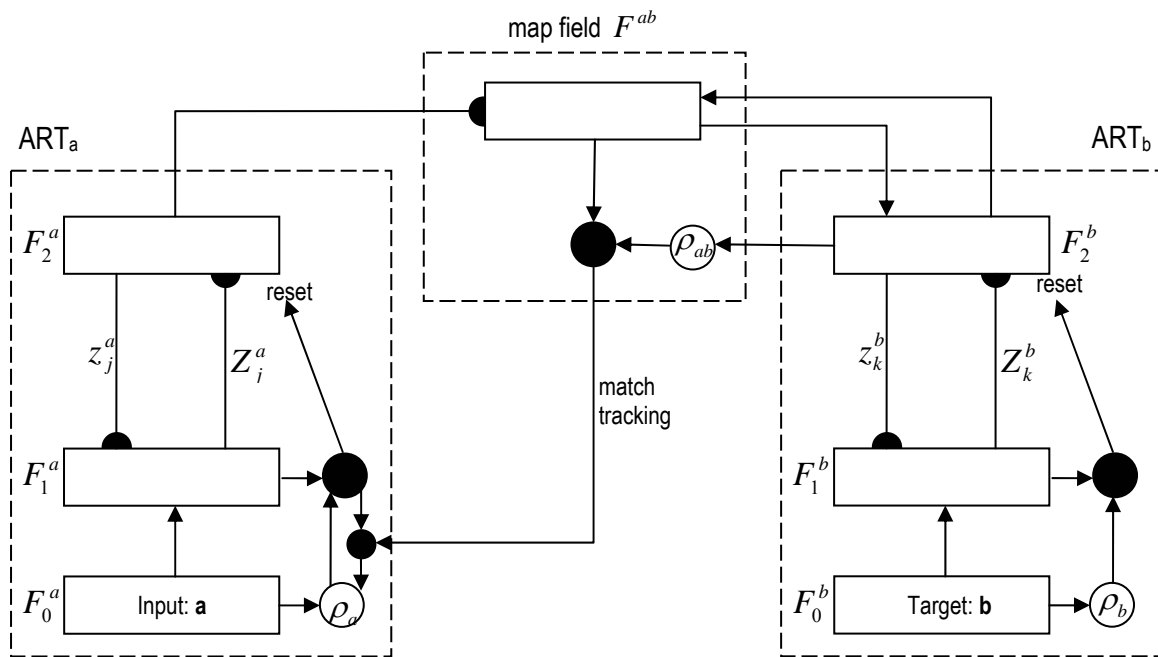
is calculated, where  $Z_{ij}$  is the strength of the bottom-up connection between  $F_1$  node  $i$  and  $F_2$  node  $j$ . Since both the input and the bottom-up weight vectors are binary with  $Z_{ij}$  being the normalized version of  $z_{ij}$ ,  $T_j$ , can also be expressed as

$$T_j = |x \cap Z_j| = \frac{|x \cap z_j|}{\beta + |z_j|} \quad (1)$$

where  $|\cdot|$  is the norm operator ( $|x| \equiv \sum_{i=1}^m x_i$ ),  $z_j$  is the binary top-down template (or prototype) of category  $j$ , and  $\beta > 0$  is the choice parameter. Then the  $F_2$  node  $J$  that has the highest bottom-up activation is selected, i.e.  $T_j = \max\{T_j / j=1,2,\dots,n\}$ . The prototype vector of the winning node  $J$  ( $z_J; z_{Ji} \in \{0,1\}, i=1,2,\dots,m$ ) is then sent down to the  $F_1$  layer through the top-down connections, where it is compared to the current input pattern: the strength of the match is given by

$$\frac{|x \cap z_J|}{|x|},$$

which is compared with a system parameter  $\rho$  called vigilance ( $0 < \rho \leq 1$ ). If the input matches sufficiently, i.e., the match strength  $\geq \rho$ , then it is assigned to  $F_2$  node  $J$  and both the bottom-up and top-down connections are adjusted for this node. If the stored prototype  $z_J$  does not match the input sufficiently (match strength  $< \rho$ ), the winning  $F_2$  node  $J$  is reset for the period of presentation of the current input. Then another  $F_2$  node (or category) will be selected, whose prototype will be matched against the input. This "hypothesis-testing" cycle is repeated until the network either finds a stored category whose prototype matches the input closely enough, or allocates a new  $F_2$  node. Then learning takes place as described above. After an initial period of self-stabilization, the network will directly (i.e., without search) access the prototype of one of the categories it has found in a given training set. The higher the vigilance level, the larger number of smaller, or more specific, categories will be created. If  $\rho = 1$ , the network will learn every unique input perfectly with a different category.



**Figure 2.** Architecture of ARTMAP network.

The architecture of the ARTMAP network can be seen in Figure 2. It consists of two ART modules that are linked together through an inter-ART associative memory, called map field  $F^{ab}$ . Module ART<sub>a</sub> (with a baseline vigilance  $\bar{\rho}_a$ ) learns to categories input patterns presented at layer  $F_0^a$ , while module ART<sub>b</sub> with vigilance  $\rho_b$  develops categories of target patterns presented at layer  $F_0^b$ . Modules  $F_2^a$  and  $F^{ab}$  are fully connected via associative links whose strengths are adjusted through learning. There are one-to-one, two-way, and non-modifiable connections between nodes in the  $F^{ab}$  and  $F_2^b$  layers, i.e., each  $F_2^b$  node is connected to its corresponding  $F^{ab}$  node, and vice versa. A new association between an ART<sub>a</sub> category  $J$  and an ART<sub>b</sub> category  $K$  is learned by setting the corresponding  $F_2^a \rightarrow F^{ab}$  link to one and all other links from the same ART<sub>a</sub> node to zero. When an input pattern is presented to the network, the  $F^{ab}$  layer will receive inputs from both the ART<sub>a</sub> module through the previously learned  $J \rightarrow K$  associative link and the ART<sub>b</sub> module from the active  $F_2^b$  category node. If the two  $F^{ab}$  inputs match, i.e., the network's prediction is confirmed by the selected

target category, the network will learn by modifying the prototypes of the chosen ART<sub>a</sub> and ART<sub>b</sub> categories according to the ART learning equations shown above. If there is a mismatch at the  $F^{ab}$  layer, a map field reset signal will be generated, and a process called match tracking will start, whereby the baseline vigilance level of the ART<sub>a</sub> module will be raised by the minimal amount needed to cause mismatch with the current ART<sub>a</sub> input at the  $F_1^a$  layer. This will subsequently trigger a search for another ART<sub>a</sub> category, whose prediction will be matched against the current ART<sub>b</sub> category at the  $F^{ab}$  layer again. This process continues until the network either finds an ART<sub>a</sub> category that predicts the category of the current target correctly, or creates a new  $F_2^a$  node and a corresponding link in the map field, which will learn the current input/target pair correctly. The ART<sub>a</sub> vigilance is then allowed to return to its resting level  $\bar{\rho}_a$ .

After a few presentations of the entire training set, the network will self-stabilize, and will read out the expected output for each input without search.

---

### ARTMAP Learning

---

All ART1 learning is gated by  $F_2$  activity - that is - the adaptive weights  $z_{ji}$  and  $Z_{iJ}$  can change only when the  $J$ -th  $F_2$  node is active. Then both  $F_2 \rightarrow F_1$  and  $F_1 \rightarrow F_2$  weights are functions of the  $F_1$  vector  $x$ , as follows:

#### Top-down learning

Stated as a differential equation, this learning rule is [3]

$$\frac{d}{dt} z_{ji} = y_j (x_i - z_{ji}) \quad (2)$$

In equation (2), learning by  $z_{ji}$  is gated by  $y_j$ . When the  $y_j$  gate opens - that is when  $y_j > 0$  - then learning begins and  $z_{ji}$  is attracted to  $x_i$ . In vector terms, if  $y_j > 0$ , then  $z_J$  approaches  $x$ . Initially all  $z_{ji}$  are maximal:  $z_{ji}(0) = 1$ . Thus with fast learning, the top-down weight vector  $z_J$  is a binary vector at the start and end of each input presentation.  $F_1$  activity vector can be described as

$$x = \begin{cases} I & \text{if } F_2 \text{ is inactive} \\ I \cap z_J & \text{if the } J^{\text{th}} F_2 \text{ node is inactive} \end{cases} \quad (3)$$

When node  $J$  is active, learning causes

$$z_J(\text{new}) = I \cap z_J(\text{old}) \quad (4)$$

where  $z_J(\text{old})$  denotes  $z_J$  at the start of the input presentation.

#### Bottom-up learning

In simulations it is convenient to assign initial values to the bottom-up  $F_1 \rightarrow F_2$  adaptive weights  $Z_{ij}$  in such a way that  $F_2$  nodes first become active in the order  $j=1, 2, \dots$ . This can be accomplished by letting  $Z_{ij}(0) = \alpha_j$ , where  $\alpha_1 > \alpha_2 > \dots > \alpha_N$ . Like the top-down weight vector  $z_J$ , the bottom-up  $F_1 \rightarrow F_2$  weight vector  $Z_J \equiv (Z_{1J}, Z_{2J}, \dots, Z_{iJ}, \dots, Z_{MJ})$  also becomes proportional to the  $F_1$  output vector  $x$  when the  $F_2$  node  $J$  is active. In addition, however, the bottom-up weights are scaled inversely to  $|x|$ , so that

$$Z_{iJ} \rightarrow \frac{x_i}{\beta + |x|} \quad \text{where } \beta > 0.$$

This  $F_1 \rightarrow F_2$  learning realizes a type of competition among the weights  $z_J$  adjacent to a given  $F_2$  node  $J$ . This competitive computation could alternatively be transferred to the  $F_1$  field, as it is in ART2 [2]. During learning

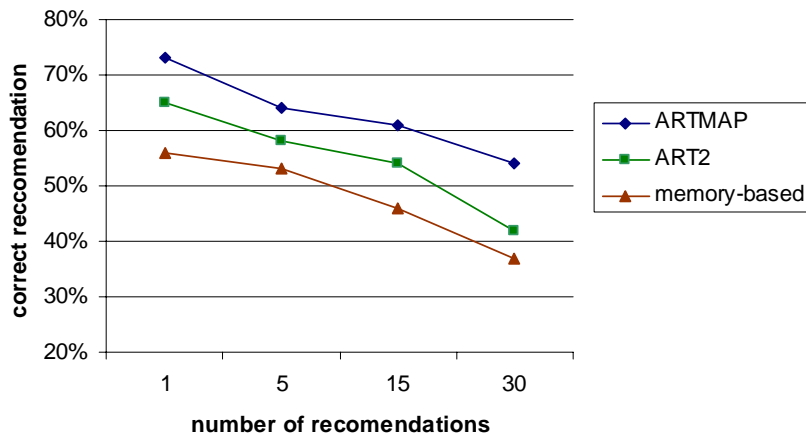
$$Z_J(\text{new}) = \frac{I \cap z_J(\text{old})}{\beta + |I \cap z_J(\text{old})|} \quad (5)$$

The  $Z_{ij}$  initial values are required to be small enough so that  $0 < \alpha_j = Z_{ij}(0) < \frac{1}{\beta + |I|}$  for all  $F_0 \rightarrow F_1$  inputs  $I$ .

## Experiments

A series of experiments were conducted to estimate ARTMAP architecture as a model-based approach to CF. For experiments a CF component, based on ARTMAP neural network was used. It was designed with 60  $F_2^a$  neurons, 40  $F_2^b$  neurons, and 40  $F^{ab}$  map-field neurons. Two other CF components were also used – one based on ART2 network with 60  $F_2$  neurons and one memory-based CF component that incorporates the popular neighbourhood-based algorithm, as described in [4].

Most of the results presented here were obtained by using the publicly available EachMovie dataset [12]. It contains 2,811,983 ratings on a scale from 1 to 5 for 1,628 movies by 72,916 users. On average, each user rated about 46.3 movies. As in [4], analysis was restricted to the users who have minimum the average for the database rating activity (45 entries) in their profile, and extracted 196817 vote records of the first 2000 of those users from the database. Restricted number of user reveals the performance of the model-based CF approach under conditions where the ratio of users to items is low. This is condition that every CF service has to go through in its first phase.



**Figure 3.** Correct recommendations with growing dataset.

The resulting dataset of users and their votes was divided into two data sets - a training set that contains randomly selected 60 rated items, and a test set with randomly selected 40 rated items. To simulate a growing database, three experiments were conducted using 30%, 60% and 100% of available profile entries, with 40 control set entries in each case that we used to evaluate the computed recommendations. The three different subsets have been used as training sets for the neural networks and as input for the memory based method. Afterwards 1, 5, 15, and 30 recommendations were computed and compared to the control set of 40 profile entries. Figure 3 summarizes the results of those experiments. It can be seen that in terms of correct recommendations in conditions of growing dataset the ARTMAP network performed better than both ART2 network and memory based method.

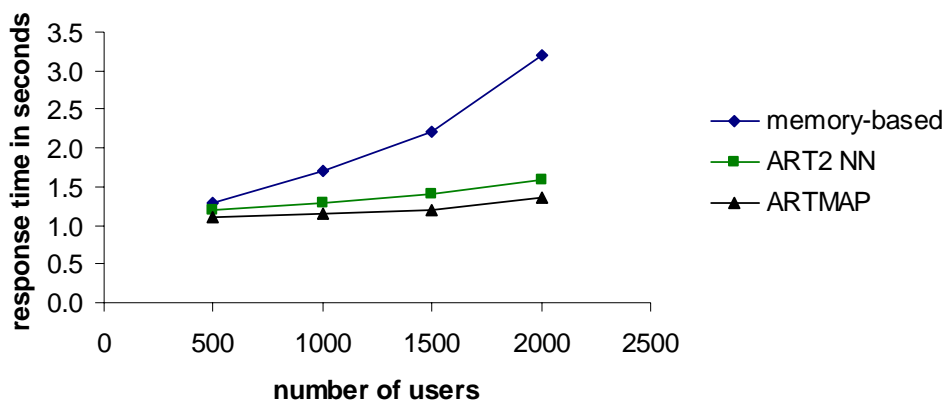


Figure 4. Response time.

Second group of experiments aimed to compare response time of both the ART2 NN and memory-based neighborhood algorithm. Five series of experiments were conducted with growing number of users. The four test sets contain profile entries for 500, 1000, 1500 and 2000 of the user data set. Each time recommendations were computed, the response time has been measured. Results summarized in Figure 4 show that the proposed ARTMAP CF component performs better than both ART2 and model-based components in terms of response time when the number of users increases. As expected and shown in Figure 4, the number of users has a much less significant influence to the performance of the neural network based methods than the memory-based one.

## Conclusion

Generally, the task in collaborative filtering is to predict the votes of a particular user from a database of user votes from a sample or population of other users. This paper presents a model based-approach to collaborative filtering by using supervised ARTMAP neural network (NN). Proposed algorithm is based on formation of reference vectors that make a CF system able to classify user profile patterns into classes of similar profiles, which forms the basis of a recommendation system. Experimental results presented here used the EachMovie data set. The first group of experiments shows classification accuracy in condition of growing database of votes. It can be seen the ARTMAP network provides better performance than both ART2 network and the popular memory-based neighborhood algorithm. The second group of experiments shows the advantage of the proposed ARTMAP model over both ART2 model and the memory-based method comparing response times in condition of growing number of users.

## Bibliography

- [1] Carpenter, G., S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 54-115, 1987.
- [2] Carpenter, G., & Grossberg, S. ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, 26, 4919-4930, 1987.
- [3] Carpenter, G., S. Grossberg, and J. H. Reynolds, ARTMAP: Supervised real-time learning and classification of non-stationary data by a self-organizing neural network, *Neural Networks*, vol. 4, pp. 565-588, 1991.
- [4] Nachev, A., I. Ganchev, A Model-Based Approach to Collaborative Filtering by Neural Networks, In proceedings of the 2004 International Conference in Computer Science and Computer Engineering IC-AI'05, Las Vegas, 2005
- [5] Goldberg, D., D. Nichols, B. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, *Communications of the ACM* 35, 1992
- [6] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, GroupLens: An open architecture for collaborative filtering of netnews, In Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175-186, 1994.
- [7] Shardanand, U., P. Maes, Social information filtering: Algorithms for automating "word of mouth." In Proceedings of Computer Human Interaction, pp. 210-217, 1995.

- [8] Breese, J., D. Heckerman and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pp. 43–52, 1998.
  - [9] Basu, C. Hirsh, H. and Cohen, W. Recommendation as classification: Using social and content-based information in recommendation, In Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), pp.714–720, 1998
  - [10] Billsus, D., and M. Pazzani, Learning collaborative information filters, In Proceedings of the Fifteenth International Conference on Machine Learning, pp. 46–54, July 1998.
  - [11] Schafer, J., J. Konstan, and J. Riedl. Recommender systems in e-commerce. In Proceedings of the ACM Conference on Electronic Commerce (EC-99), pp. 158–166, 1999.
  - [12] Mc Jones, P., EachMovie collaborative filtering data set, DEC Systems Research Center, <http://www.research.compaq.com/SRC/eachmovie/>.
  - [13] Hawes, D. Information literacy in the business schools, In: Sep/Oct Journal of Education in Business, 70, pp 54-62, 1994
- 

### Author's Information

---

Anatoli Nachev – Information Systems, Dept. of A&F, NUI Galway, Ireland; e-mail: [anatoli.nachev@nuigalway.ie](mailto:anatoli.nachev@nuigalway.ie)

## SYNTHESIS METHODS OF MULTIPLE-VALUED STRUCTURES OF LANGUAGE SYSTEMS

Mikhail Bondarenko, Grigorij Chetverikov, Alexandr Karpukhin,  
Svetlana Roshka, Zhanna Deyneko

**Abstract:** *The basic construction concepts of many-valued intellectual systems, which are adequate to primal problems of person activity and using hybrid tools with many-valued coding are considered. The many-valued intellectual systems being two-place, but simulating neuron processes of space coding which are different on a level of actions, inertial and threshold of properties of neurons diaphragms, and also modification of frequency of following of the transmitted messages are created. All enumerated properties and functions in point of fact are essential not only are discrete on time, but also many-valued.*

**Keywords:** *intelligence system, hybrid logic, multiple-valued logic, multi-state element.*

---

### Introduction

---

The basic construction concepts of many-valued intellectual systems (MIS), which are adequate to primal problems of person activity and using hybrid tools with many-valued coding [1, 2] are considered. With materialism of a point of view these concepts are agreed with the dialectic laws opened by a man and their manifestations in problems connected with creation of identification systems prediction and recognition of imagery in which the interactive operational mode is a main part of the whole complex of intellectual properties.

Those are, for example, the law of unity and struggle of contrasts – as availability in parallel operating in space and time of mechanisms both discrete, and continuous mapping objects of plants; the law of transition from quantitative changes to qualitative-quantitative changes of gradation levels of brightness and the colors result in qualitative changes in mapping of objects; the law of negation of negation – as a changes and alternation of coding indications of messages about objects in neurons of a brain – from space to temporal and from two-place to many-valued [3,5].

In particular, in works the accent on the concept of neuro-physiologic and neuro-cybernetic aspects of alive brain mechanisms is made. It is connected with the following natural neuron structures from nervous cells – neurons, essentially are highly effective recognizing systems and, for this reason, is of interest not only for doctors and physiologists, but also for the experts designing artificial intelligence systems. However direct transfer of research results of neuro-physiologists in engineering practice is now impossible because of a lack of an appropriate

bioelectronic technology and an element basis, that has led to development and creation of a set of varieties of artificial neurons realized on the elements of the impulse technology.

But also here there were complications because of non-adequate neuron models to a set of the demands made of MIS. Creation of neuro-like models on the basis of multiprocessor in inputting systems technology with programmed architecture, in particular, on the basis of digital integrating structures is offered as the alternative in works [1–4]. Thus, retaining Neumann structure a MIS are created, being essentially two-place, but simulating neuron processes of space totting different on a level of actions, inertial and threshold properties of neuron diaphragms, as well as variation of recurrence frequency of transmitted messages. Though it is obvious that all enumerated properties and functions in point of fact, are, essential, not only discrete on time, but also many-valued (are discrete on a level).

As the corollary, non-adequacy of used principles of coding and element basis to simulated processes entails a redundancy, complication and non evidence of used mathematical and engineering means of transformations, loss of a micro level of parallelism in handling expected fast acting and flexibility of restructuring without essential modifications of architecture and connections.

### Structurally Functional Cell Model of a Many-Valued Intellectual System

The originating complications [1], in creation of a many-valued intellectual system (MIS) promote moving out of the adequacy concept of many-valued logic and structures to of MIS creation problems with desirable properties and possibilities.

Therefore, for disclosure of use paths of a knowledge backlog in the field of many-valued coding and structures in MIS creation the conceptual structurally functional model of a MIS cell (Fig.1) is offered.

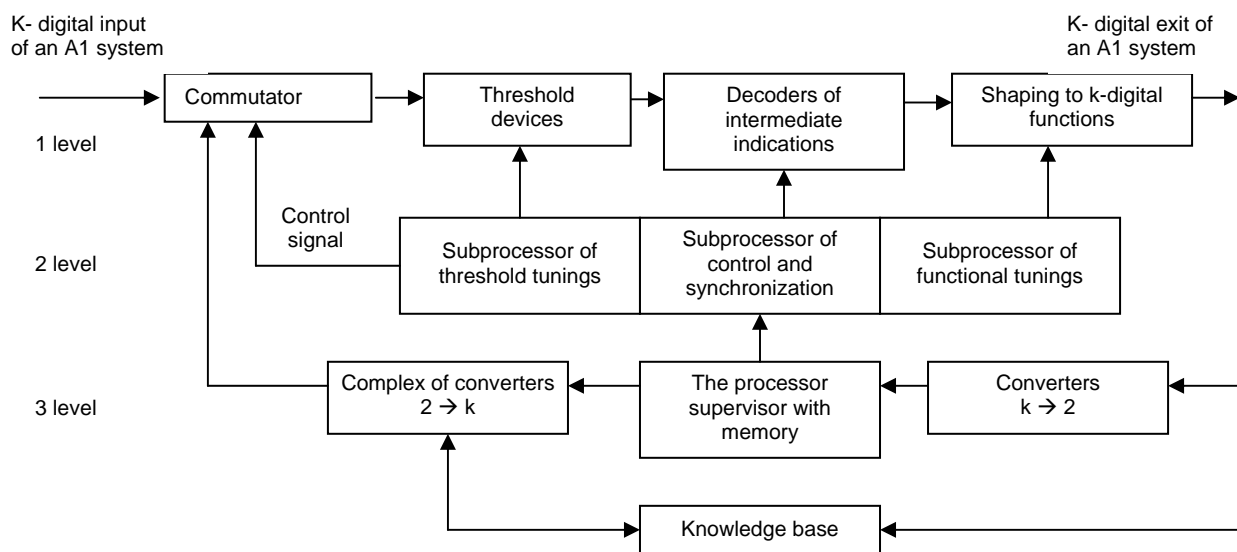


Fig. 1. A conceptual structurally functional model of a MIS cell.

Each MIS is characterized by a set of functions fulfilled by it by blocks, which realize functions and information interchanges. In accordance with solved problems, the structurally functional cell breaks up to three hierarchical levels: functional (analytic-synthetic) – level 1; tactical (analyses-coordination) – level 2; strategic (coordination) – level 3.

The MIS cell increases on a function level both on inputs, and on outputs, and it is integrated with other meshes on inputs of decoders of intermediate indications; at a tactical level – through the analyze-coordination processor; at a strategic level – through the processor-supervisor and knowledge base. The conceptual model of a MIS cell

is based on the concept of symbiosis of two- and many-valued tools of data processing, therefore at a strategic level it contain complexes of converters of the data representation form – converters from a two-place code to many-valued (2→K) and back (K→2). Obviously, that their use in MIS determines, at what level the problems, are solved in what logic and with what speed (what channel capacity of MIS). Besides the application of these tools excludes necessity of an operator work with two-place translators in input – output of data.

The new principle of the COMPUTER construction is offered, in which the principle of organization of brainwork simultaneously with a principle of programmed control assumes as a basis. The principle of organization of brainwork assumes as a basis of operation of such COMPUTERS, in classical element basis it will be for more to Hilbert machines than for nowadays existing Neumann machines, the basis of which is the principle of programmed control realized rather slowly.

**Formalization of Construction Principles of Many-Valued Spatial Structures**

In the generalized form the two-input universal k-valued structure of a spatial type contains two recognition elements (RE), the control unit (CU), the matrix selector (MS), commutator (C), and keys (K) or the digital-to-analog converter (DAC) [2,3] (Fig.2).

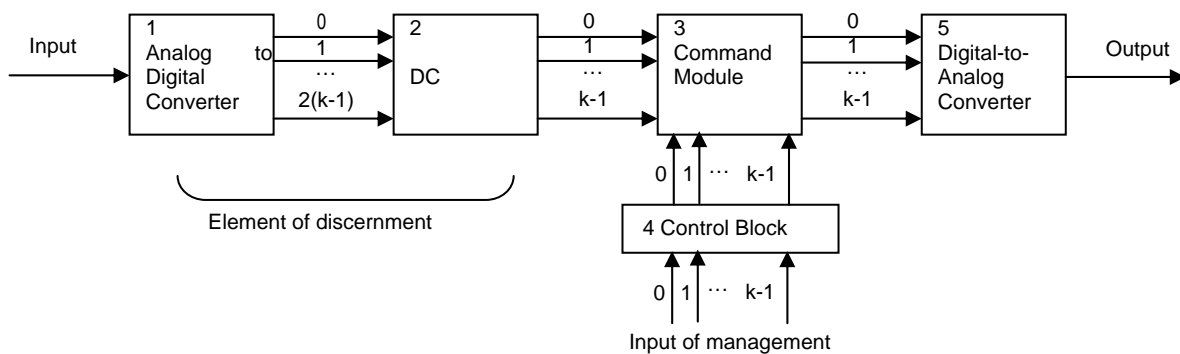


Fig. 2. Universal Multiple-Valued Functional Converter.

The logic of the decoders operation in recognition elements 1,2 is described by the following equation system:

$$\begin{aligned}
 f_0 &= (x_0, x_1, \dots, x_{k-1}) = y^0, \\
 f_1 &= (x_0, x_1, \dots, x_{k-1}) = y^1, \\
 &\dots, \\
 f_{k-1} &= (x_0, x_1, \dots, x_{k-1}) = y^{k-1}.
 \end{aligned}$$

or in the explicit form at the algebra language of finite predicates [1]:

$$\begin{aligned}
 y_{1,2}^0 &= \overline{x_1}, \\
 y_{1,2}^1 &= x_1 \cup \overline{x_2}, \\
 y_{1,2}^2 &= x_2 \cup \overline{x_3}, \\
 &\dots, \\
 y_{1,2}^{k-1} &= x_{k-1}.
 \end{aligned}$$

where  $x_i$  and  $\overline{x_i}$  ( $i = 0, k-1$ ) – signals of direct and inversion outputs of the ADC units in recognition elements 1,2. The logic of the matrix selector is described by the following equation system:



$$\begin{aligned}
b_{00} &= y_1^0 \cup y_2^0, b_{01} = y_1^0 \cup y_2^1, \dots, b_{0(k-1)} = y_1^0 \cup y_2^{k-1} \\
b_{10} &= y_1^1 \cup y_2^0, b_{11} = y_1^1 \cup y_2^1, \dots, b_{1(k-1)} = y_1^1 \cup y_2^{k-1} \\
&\dots\dots\dots \\
b_{(k-1),0} &= y_1^{k-1} \cup y_2^0; b_{(k-1),1} = y_1^{k-1} \cup y_2^1; \dots b_{(k-1),(k-1)} = y_1^{k-1} \cup y_2^{k-1}
\end{aligned}$$

where  $b_{ij}$  ( $i, j = \overline{0, k-1}$ ) – output logical signals of the matrix selector<sup>4</sup>. The commutator has two groups by  $k$  inputs: the signals from the selector are applied to the first group and control signal values are, applied to the second group. In the explicit from the commutator operation is described by the following system:

$$\begin{aligned}
b^{k_0} l^0 \cup b^{k_0} l^1 \cup \dots \cup b^{k_0} l^{k-1} &= z^{k_0}, \\
b^{k_1} l^0 \cup b^{k_1} l^1 \cup \dots \cup b^{k_1} l^{k-1} &= z^{k_1}, \\
&\dots\dots\dots \\
b^{k_{k-1}} l^0 \cup b^{k_{k-1}} l^1 \cup \dots \cup b^{k_{k-1}} l^{k-1} &= z^{k_{k-1}}.
\end{aligned}$$

As all  $k$  of keys of the output shaper are constantly connected to corresponding  $k$ -values of output signals the function values selected by the commutator and the control unit, respectively, will arrive in the converter output (structure) in the course of variations of  $k$ -valued functions on the converter inputs. The process control of the logic recombinations is carried out under the action of external control signals.

## Modeling and Realization

One of ways of realization of multiple-valued elements is the frequent-harmonic multi-stable element, which basis is the self-excited oscillator with a nonlinear resonant circuit, which is synchronized by an external voltage source.

At apparent simplicity, such circuit due to nonlinear properties has a lot of stable states. This circuit is supplied by a sequence of pulses with high period-to-pulse duration ratio. The control of circuit is carried out by feed of control pulses in a circuit of automatic bias. The process comes to an end then, when the resonant circuit appears tuned on next harmonic of a supplied voltage. Besides, the voltage of automatic bias changes too. Thus, the multi-state element has two attribute of each stable state – a voltage and frequency.

Using parabolic approximation of the characteristic of the transistor, we shall receive the following equation for a charge on nonlinear capacity of MOS-structure.

$$\begin{aligned}
\frac{d^2 \chi}{dt^2} + \omega^2 \chi &= -F_2 \left( \frac{d^2 \chi}{dt^2}, \frac{d\chi}{dt}, \chi \right) + S^*(t); \\
F_2 \left( \frac{d^2 \chi}{dt^2}, \frac{d\chi}{dt}, \chi \right) &= \varepsilon \frac{d^2 \chi}{dt^2} (\chi^2 \alpha + \chi(1 + \beta) + \gamma) + \varepsilon \left( \frac{d\chi}{dt} \right)^2 \cdot (1 + \chi(1 + \beta) + \gamma) + \\
&+ \frac{d\chi}{dt} \left( h_1 + \frac{\varepsilon}{\tau} (1 + \chi^2 \alpha + \chi(1 + \beta) + \gamma) \right) + \omega^2 \left( 1 + \frac{\chi^3}{3} + \chi^2 \alpha + \chi \beta + \gamma - B^* \right)
\end{aligned}$$

where  $\chi$  – normalized charge on capacity of MOS-structure;  $\omega$  – resonant frequency of a resonant circuit;  $\varepsilon$  – small parameter;  $S^*(t) = N_k P(\tau_1) \sin[k\Omega, t + \gamma_k(\tau_1)]$

$$C_k = C_{k0} (1 + b)^{\frac{1}{2}}; s_0^* = \frac{s_1}{C_k}; s_2^* = s_0^* + \frac{\xi}{\varepsilon}; \lambda = s_1^* \varphi_k (1 + b);$$

$s_0, s_1$  – coefficients of polynomial in approximation of the characteristic of the transistor;

$\varphi_k$  – contact difference of potentials.

The solution of the this equation in the second approximation in case of the main resonance is

$$\begin{aligned} \chi &= \alpha \cos \psi; \quad \psi = \nu t + \mathcal{G}; \\ \frac{d\alpha}{dt} &= \alpha \xi - \alpha^2 \delta - \alpha \eta - N_k \frac{P(\tau_1)}{\omega + \nu} \cos[\gamma_k(\tau_1) - \mathcal{G}]; \\ \frac{d\mathcal{G}}{dt} &= \omega - \nu + \chi + \frac{1}{\alpha} \xi + \alpha \theta + \alpha^2 \sigma + N_k \frac{P(\tau_1)}{\alpha(\omega + \nu)} \sin[\gamma_k(\tau_1) - \mathcal{G}], \end{aligned} \quad (1)$$

where:

$\nu$  – frequency of a synchronizing signal;

$$\begin{aligned} \xi &= \frac{1}{\pi} \left( \left[ h_1 + \frac{\varepsilon}{\tau_y} (1 + \gamma) \right] \left( \frac{\sin 2\psi_1}{4} - \frac{\psi_1}{2} \right) - \left( h_1 + \frac{\varepsilon}{\tau_e} \right) \left( \frac{\sin 2\psi_1}{4} + \frac{\pi}{2} - \frac{\psi_1}{2} \right) \right); \\ \theta &= \frac{1}{\pi\omega} \left[ H \left( \frac{\sin 3\psi_1}{12} + \frac{3}{4} \sin \psi_1 \right) + \frac{1}{3} \varepsilon \omega^2 (1 + \gamma) \sin^3 \psi_1 \right]; \\ H &= \omega^2 \alpha - \varepsilon \nu^2 (1 + \beta). \end{aligned}$$

Let's consider conditions, at which the stable synchronous mode of stationary oscillations is possible. The values of amplitude and phase in a stationary mode are determined from system of the equations

$$R(a, \nu) = 0; \quad T(a, \nu) = 0, \quad (2)$$

where  $R(a, \nu), T(a, \nu)$  – right parts of the equations (1).

The conditions of stability of the solutions of the equations (2) are determined by the following inequalities:

$$\begin{aligned} \frac{da}{d\nu} &> 0 \quad \omega_e(a) > \nu; \\ \frac{da}{d\nu} &< 0 \quad \omega_e(a) < \nu, \end{aligned}$$

where  $\omega_e(a)$  – equivalent frequency of own oscillations.

The analysis of phase portraits of this dynamic system (frequency-harmonic multi-state element) has confirmed presence of stable modes in it.

By excluding a phase from the equations (2) it is possible to receive the equation for the amplitude-frequency characteristic [6].

The considered above frequency-harmonic multi-state element was realized as the hybrid thin-film integrated circuit with MOS-structure chip. Inductance elements were made as thin film LC structure [7]. The problem of it optimization [8] was solved.

The earlier received results get the importance in this time, when the semiconductor technology of manufacturing of the large scale integrated circuits for microprocessors practically has reached a physical limit of reduction of the size of components and width of interconnections. Alternative can be only use of artificial language systems, in which the elements of multiple-valued logic can be used. Experimental samples of the frequency-harmonic multi-state element were realized as thin-film integrated circuits in the standard case and can be used as elements of multiple-valued logic.

---

## Conclusion

---

The problem solving of principles formalization of the structure organization of computing tools, thus ensures construction of the newest concept for systems of an artificial intelligence; application of space and temporal parallelism at structural and algorithmic levels; creation of procedural and functional languages, parallel machines of knowledge bases and the inference. The problem solving of organization principles formalization of universal k-valued structures of a spatial type by tools of predicate and hybrid logic will ensure construction of a modern concept for artificial intelligence systems, application of spatial parallelism at structured and algorithmic levels; creation of functional languages of parallel machines of knowledge basis; application of symbiosis of two- and many-level heterogeneous coding.

One of circuit for realization of multiple-valued elements is the frequency-harmonic multi-state element which states are coding by amplitude and frequency. This element was made by thin film technology as hybrid integrated circuit.

---

## Bibliography

---

- [1] *M.F. Bondarenko, Z.D. Konopljanko, G.G. Chetverikov. Osnovy teorii synteza nadshvydkodiuchikh struktur movnykh sistem shtuchnogo intelektu, Monografia. – K.: IZMN, 1997. – 386 s.*
- [2] *M.F. Bondarenko, Z.D. Konopljanko, G.G. Chetverikov. Osnovy teorii bagatoznachnikh struktur i koduvannya v sistemach shtuchnogo intelektu. – Kh.: Factor-Druk, 2003. – 336 s.*
- [3] *M.F. Bondarenko, S.V. Lyahovets, A.V. Karpukhin, G.G. Chetverikov. Sintez shvidkodiuchikh struktur lingvistichnich ob'ekhtiv., Proc. of the 9<sup>th</sup> International Conference KDS – 2001, St.Peterburg, Russia, 2001, s. 121–129.*
- [4] *M.F. Bondarenko, A.V. Karpukhin, G.G. Chetverikov. Analiz problemi sozdaniya novich tekhnicheskikh sredstv dlya realizazii lingvisticheskogo itterfeisa.Proc. of the 10<sup>th</sup> International Conference KDS – 2003, Varna, Bulgaria, June16–26, 2003, pp. 78-92.*
- [5] *M.F. Bondarenko, V.N. Bavykin, I.A. Revenchuk, G.G. Chetverikov. Modeling of universal multiple-valued structures of artificial intelligence systems, Proc. of the 6<sup>th</sup> International Workshop “MIXDES'99”, Krakow, Poland, 17-19 June 1999, pp. 131–133.*
- [6] *M.F. Bondarenko, A.V. Karpukhin, G.G. Chetverikov, Zh.V. Deyneko. Application of a numerically – analytical method for simulation of non-linear resonant circuits.10 th International Conference Mixed Design Of Integrated Circuits And System (MIXDES 2003), Lodz, Poland, 26–28 June 2003,*
- [7] *V.V. Aleksandrov, A.V. Karpukhin. Osobennosti konstruktivnogo rascheta i tehnologij izgotovlenija mikroelektronnich ustrojstv obmena informaciej. Izvestija visshich uchebnich zavedenij.Priborostroenije. Leningradskij institut tochnoj mehaniki i optiki. Tom. XX, N 5, 1977. – pp.120-124.*
- [8] *G.I. Yalovega., Yu.Kh. Loza, A.V. Karpukhin. Matematicheskoe modelirovanie i optimizaciya mnogofunkcionalnich resonansnikh cepei. 26. Internationales Wissenschaftliches Kolloquim.Technische Hochschule Ilmenau, 1981, Heft 2, Vortragsreihen A3, A1.*

---

## Authors' Information

---

**Mikhail Fedorovich Bondarenko** – Rector, Prof., State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine.

**Grigoriy Grigorjevich Chetverikov** – c.t.s. State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine.

**Alexandr Vladimirovich Karpukhin** – c.t.s., State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine; e-mail: [kav@kture.kharkov.ua](mailto:kav@kture.kharkov.ua)

**Svetlana Alexandrovna Roshka** – Ph.D. student, State National University of Radio-Electronics P.O. Box: 14, Lenin's avenue, Kharkov, 61166, Ukraine.

**Zhanna Valentinovna Deyneko** – Ph.D. student, State National University of Radio-Electronics P.O. Box: 14, Lenin's avenue, Kharkov, 61166, Ukraine.

## ARCHITECTURE AND PRINCIPLES OF CONTROL FOR MULTI-AGENT TELECOMMUNICATIONAL SYSTEMS OF NEW GENERATION

**Adil V. Timofeev**

Global multi-agent telecommunication systems (TCS) serve for providing to users informational and computing resources, distributed in computer networks (CN). Architecture of such TCN consists of four main subsystems:

- distributed communication system (DCS);
- network control system (NCS);
- distributed informational system (DIS);
- distributed transport system (DTS).

All mentioned systems of multi-agent TCS have distributed property, interconnection and interact actively between each other in a process of providing for users the informational and computing resources, storing in global CS.

The main role in aiming and quality information processing and address transfer of data flows by users queries is played by NCS. It obtains through DCS client queries and commands of network administrators of TCS, processes internal information about current state of DTS and external information about state of information and computing resources in CS, coming from DIS, and forms DTS control, providing satisfaction of user queries by way of transfer to them necessary informational and computing resources of CS.

The main task for NCS of global TCS of new generation, working on a big speed of data flows, is adaptive forming of multi-agent control for traffic of heterogeneous data of big volume with reliable guarantees of high quality of service (Quality of Service, QoS) of TCS users. Solution of this task is divided on local tasks for control of data flows, adaptation to changing traffic, overloading avoidance, resolution of network conflicts etc.

Traditionally for organization of control of data flows and equipment of DTS network principles and architectures of centralized and decentralized control are used. With consideration of disadvantages of traditional network architectures, it is useful to develop "hybrid" architecture of NCS for global TCS of new generation, combining in itself advantages of centralized and decentralized architectures. Let name this "compromise" architecture a multi-agent architecture of NCS.

In this case the main functions of information processing and control for data flows in global TCS of new generation is distributed between interconnected intelligent agents. Every network agent has own local DB or KB and tools for communication with other agents for information exchange in process of joint (cooperative) solution making and automated forming of network control for DTS, providing address delivery of informational computing resources of KS by global TCS users queries.

Network agents may be executed by NCS computers, connected with DTS nodes and also software agents of DCS and DIS. Let name such agents internal agents of global TCS. Then role of external agents will be played by users (clients, administrators etc.) together with access tools in TCS and network interface and also computer nodes (hosts) of CS.

In process of design of NCS on the base of theory of agents new problems for organization of multi-address and multi-flow addressing and multi-agent dialogue between internal agents of global TCS of new generation and external agents-users and agent of CS as distributed data store and application arise. To solve these problems it is necessary to develop methods of avoidance and automated resolution of network conflicts under control of adaptive and intelligent NCS with multi-agent architecture.

For controlled address transfer and navigation of data flows, resolution of network conflicts, functional diagnosis and recognition of states of global TCS of new generation it is necessary to enter special agent-coordinators (for example, on level of data flows routing) and, probably, other global agents. Peculiarity of these agents of high level is that their DB and KB are formed on the base of local DB and KB of agents with lower level. Therefore they have global (multi-agent) property and allow to evaluate network situation in a whole.

Thus, development and modernization of NCS architectures of global TCS of new generation should be done not only "by width" (i.e. "by horizontal" of territory envelopment), but "by depth"(i.e. "by vertical" of evolution of hierarchy of network control). Important role in it is played by processes of adaptation and intellectualisation of NCS.

Let describe the main peculiarities of these processes on the example of multi-agent and adaptive routing of informational flows in TCS. Necessity in adaptive routing arises in unpredictable changes of structure (nodes and communication channels) of TCS and number of users or at overloading of buffers of nodes or communication channels of TCS. Really it is routing in non-stationary global TCS with changing structure and load.

Adaptive and multi-agent routing of data flows in global TCS have a series of advantages:

- workability and reliability of TCS at unpredictable changes of their structure and parameters are provided
- load of nodes and TCS communication channels by "equalizing" of load is balanced
- control for data flows is simplified and adaptation to network overloading becomes easier
- time for faultless work and productivity of TCS at high level of provided services in unpredictable conditions of network parameters and structure is increased, that is important for agents-users of TCS of new generation.

Principles of adaptive routing may be divided on three classes depending on used methods of processing of local or global information (feedback):

- centralized (hierarchical) routing;
- decentralized (distributed) routing;
- multi-agent (multi-address, multi-flow) routing.

Principle of multi-agent routing is compromise between principles of centralized and decentralized routing. It is based on multi-address and multi-flow routing and analysis of possible network conflicts for their avoidance or resolution in process of controlled transfer of packages by a set of optimal routes from nodes-sources to nodes-receivers. More detailed description of this principle and concrete methods of multi-agent and neural routing is in works [1-5].

Work is done under support of Minpromnauka of RF (project N 37.029.11.0027), RFBR (project 03-01-00224a) and RHSF (project N 03-06012019b).

---

## Bibliography

---

1. Timofeev A.V. Problems and Methods of Adaptive Control of Data Flows in Telecommunication Systems. – Informatization and Communication, N 1-2,2003, p.68-73 (In Russian).
2. Timofeev A.V. Models of Multi-Agent Dialogue and Informational Control in Global Telecommunication Networks. – Proceedings of 10-th International Conference "Knowledge-Dialogue-Solution" (June 26-26, 2003, Varna), 2003, p. 180-186 (In Russian).
3. Timofeev A.V. Methods for High quality Control, Intellectualization and Functional Diagnosis of Automated Systems. – Mechatronics, Automation,Control, 2003, N2, p. 13-17 (In Russian).
4. Syrtzev A.V., Timofeev A.V. Neural Approach in Multi-Agent Rotuing for Static Telecommunication Networks – International Journal "Information Theories and Their Applications", 2003,v.10,N2, p. 167-172.
5. Timofeev A.V. Multi-Agent Information Processing and Adaptive Control in Global Telecommunication Networks. – International Journal "Information Theories and Their Applications", 2003,v.10,N1, p. 54-80

---

## Author's Information

---

**Adil Vasilevich Timofeev** – Saint-Petersburg Institute for Informatics and Automation; 199178, 39, 14-th Line; Saint-Petersburg, Russia, e-mail: [tav@iias.spb.su](mailto:tav@iias.spb.su)

---

---

## Software Engineering

---

---

### APPLYING HIERARCHICAL MVC ARCHITECTURE TO HIGH INTERACTIVE WEB APPLICATIONS

**Micael Gallego-Carrillo, Iván García-Alcaide, Soto Montalvo-Herranz**

**Abstract:** *This paper presents a very new architecture for developing high interactive web applications. At the present time, there are many applications based on web. To manage, extend and correct them can be difficult due to the navigational paradigm they are based on. From here we would like to contribute to make these tasks easier taking advantage of experience obtained from the development of standalone applications in the past. Therefore, we would like to begin to settle the concepts and tools for a new framework.*

*There are other frameworks and APIs offering MVC architectures to web applications, but we think that they are not applying exactly the same concepts. While they keep on basing their architectures on the navigational paradigm, we are offering a new point of view based on an innovator hierarchical model. First, we present in this paper the main ideas of our proposal. Next, we expose how to implement it using different Java technologies. Finally, we make a first approach to our hierarchical MVC model. We also compare shortly our proposal with the previously cited technologies.*

**Keywords:** *Web Applications Engineering, Model, View, Controller, MVC.*

---

#### Introduction

---

When the web was created, its main goal was to provide to the scientific community the chance of sharing information through hypertext in a comfortable way. At the beginning, files supporting this information were completely static and it could only be changed modifying the content of the files manually. Nowadays, we find a great evolution at this respect and we can see really complex services offered by web sites. This way, we are able of managing complete applications based on web. This fact is more than interesting for giving the possibility of using an application almost everywhere through a network such Internet without installing complicated clients but a web browser.

This is the main reason for enterprises and other organizations to base their development interfaces in web, but this may not be so easy. Patterns and many other design tools look forward to find a way to make software products extensible, manageable, reusable and easy to support. MVC architecture appeared to reach this point for window based interfaces. Now it seems to be apprehensible to think if it would be possible to take MVC advantages to web based interfaces.

In this document we propose a MVC architecture for web applications based on Java 2 Enterprise Edition [J2EE, 2005]. First, we will describe the main concepts and organization of our proposed architecture. Although a brief glance to the used technology can be found at this point, this description would be applicable to any other development tool than Java. Next, we describe in detail the implementation of this architecture. Our main goal is to establish the basis to get a complete framework for developing high interactive web applications in a comfortable way.

---

## Description of the MVC Architecture in Web Applications

---

In this paper we present the base and main concepts to raise a framework to develop web interfaces for applications using MVC architecture [Krasner, 1988]. This architecture reduces the coupling between classes and increases their cohesion. This fact gets the code to be more independent so it can be easily reused to save time and effort in further developments. MVC has been widely implemented in graphics user interfaces to separate the entity responsible of showing information (view), the one responsible of storing it (model) and the one responsible of receiving user events (controller). The standard UI technology used in Java, Swing [Swing, 2005], also uses MVC.

As expected, the design we are writing about has three clearly separated parts, each one of them assumes a different responsibility corresponding to MVC model. They are the model, the view and the controller. As follows, we describe them:

### The Model

This component is the responsible for data maintenance. It has to keep it available for the application in a consistent and safe environment, not allowing any external or internal intervention to affect in any negative aspect. Every single code that implements the model has to be reusable by any other kind of interface without modifying it.

To implement this, we need a set of classes. In order to take advantage of many other technologies, these classes have to preserve JavaBeans specification [JavaBeans, 2005].

### The View

The view is the responsible for the application interface. In this case, it is the HTML and JavaScript code to send it to the client (or any other format like WML). It has to show and to receive the information managed by the application, so it can interact with the user. From here, it has to be given the chance of sending information to the application through events. It is considered a real event a request to the server, therefore, the view has to provide mechanisms like links and buttons to carry out the appropriate requests.

The way to implement this part is through JSP technology [JSP, 2005]. Every JavaBean of the model may have at least one JSP file to be presented. From these files it is used different technologies to access the data to be managed like expression language and tag libraries [Taglibs, 2005].

### The Controller

It is the responsible for controlling the interactions with the user. Whenever the application receives something produced by the user, the controller has to decide what to do next. This is the part that manages the global flow control of the application.

From a web browser, there is only one possibility of sending information to the server. This is done sending a request to a web resource. When requesting, the user can attach more information (for example from a form) than the reference to the new wanted resource. There should be one controller by each class that appears in the model and is presented through JSP files. The controller will be a Java common class with methods that receive data about what happened, interpret it and execute its code consequently to continue with the application.

### The Events System

In order to follow the MVC architecture, the user interacts with the application raising events. But the problem is that web pages have not the possibility of raising events more than requests. We call this kind of events real events. When an object of the model is presented it can give the chance of updating its information through a form. When the real event arrives to the controller, it carries the information from the form, so it can guess if something has happened or changed. We consider every single change between showed data to the client and received from it as an event and we reference them as deferred events pack. It is important to remark that the order or the time in which they are produced is not relevant at this point.

### Implementing MVC with J2EE

From the time that web was used for something than showing static information, developers had to solve a lot of problems derived from HTTP and HTML because they were not thought to support, so many things necessaries to give dynamism to web. So, new concepts and tools were developed to make programming web applications easier and possible such sessions, cookies, etc. We will try to take advantage of these and many other technologies when considered appropriate to help the MVC implementation.

In this point it is explained how to implement our MVC model and what technologies we are using for it. In Figure 1 we show a scheme of this and in Figure 2 a web application request collaboration diagram.

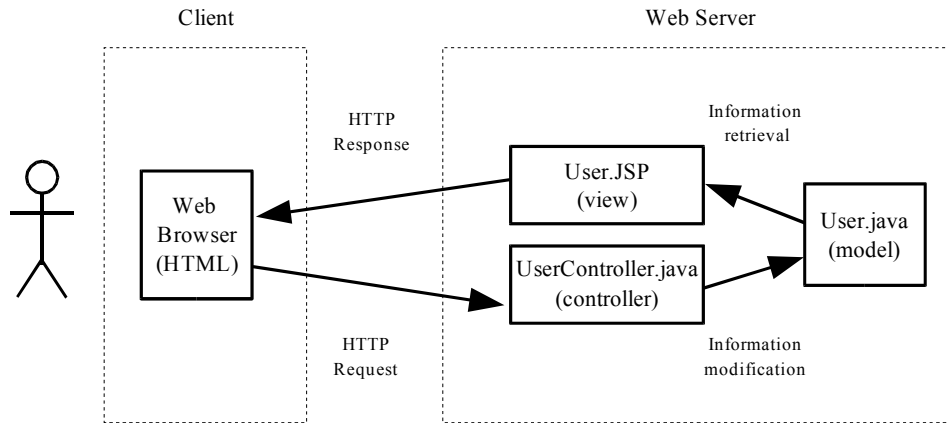


Figure 1. Scheme of the design. It shows the collaboration between different layers

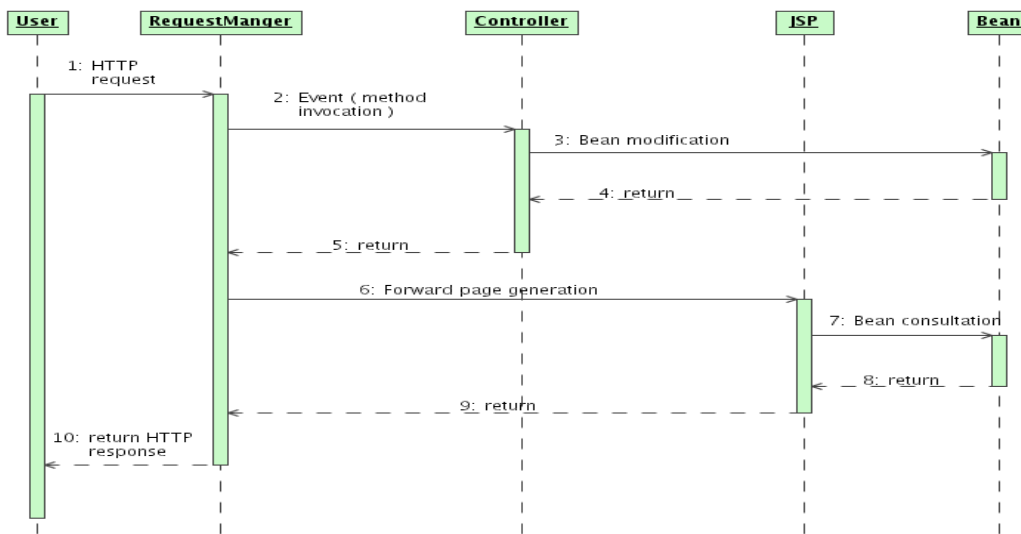


Figure 2. Web application request collaboration diagram

### Implementing the Model

In the model, we need Java classes that support the data to be managed by the application. Fulfilling JavaBeans specification will later help us to be able of using technologies that have JavaBeans as requirement. Information supported by these classes is contained in their attributes and the type of them can be another class of the same model giving rise a composition hierarchy or an association relation. Hence, basically these classes will have just some attributes and methods to access them.

In this component may appear classes related to persistent storage technologies, like data bases, if needed to get or store information in that way.



### Implementing the View

View implementation is mainly based on JSP files that access to model to show it. As classes in the model are JavaBeans, it will be more comfortable to apply Expression Language and User Defined Tag Libraries. One JSP file is always associated to a class in the model and its responsibility is to present the whole or a part of the information containing on it. One class of the model may be associated to one or more JSP files. It will be convenient to associate more than one JSP file, if it is needed, to represent the information with different formats or parts of the associated class.

Representing the application is not made from a navigational point of view what means that there is not a home page to navigate to the rest of application. Instead of, the representation of an object is sent to the user through HTTP. Because the represented object can receive events, it is its responsibility to give the appropriate links and forms to the client. If JSP file do this directly, it would be keeping the navigational paradigm, so links and form that generate real events should be generated dynamically. To get it, we let user-defined tags to take care of it. It has a code behind able of producing the text necessary for sending the request correctly to the server when the user click on a link or submit a form. Hence, we can not say that there are pages to navigate through, but there is the possibility of executing different code of the application according to the given request.

### Implementing the Controller

The controller will be a set of classes that receive information from the requests and execute the code of the application according to it. It is needed something to manage requests to the server in order to be executed them by the correct controller. This responsibility will be taken by the request manager, which is a Servlet that it is included in the framework. On the other hand, there will be a controller class by each JSP file that is able of sending requests. Because it is not necessary to keep the state of these classes, all their methods will be static. Every controller class has one method associated to each link or form that its JSP file is able to generate in order to attend it. Every method will receive one parameter with the object of the model that the JSP used to represent it. If there is additional information on the request like components of a form or parameters of a link, this information will be received by the method also in order to let the controller to make the opportune decisions.

### Hierarchical MVC Architecture

To build the component model using object oriented programming, it is often used composition hierarchies between classes. During the development of the user interface, it would be very useful to be able of using the same composition architecture. To get it, each class in the model has one or more JSP files in the view that will take the responsibility of presenting it. When the representation of one object has another object embedded by a composition relation, the JSP file associated to the first object will delegate to the corresponding JSP file associated to the related one its representation. With this approach, a JSP file may have to include another one or more if in the model one class has attributes of another model class type. It is showed in Figure 3. For implementing this approach, it is very useful to have another user-defined tag.

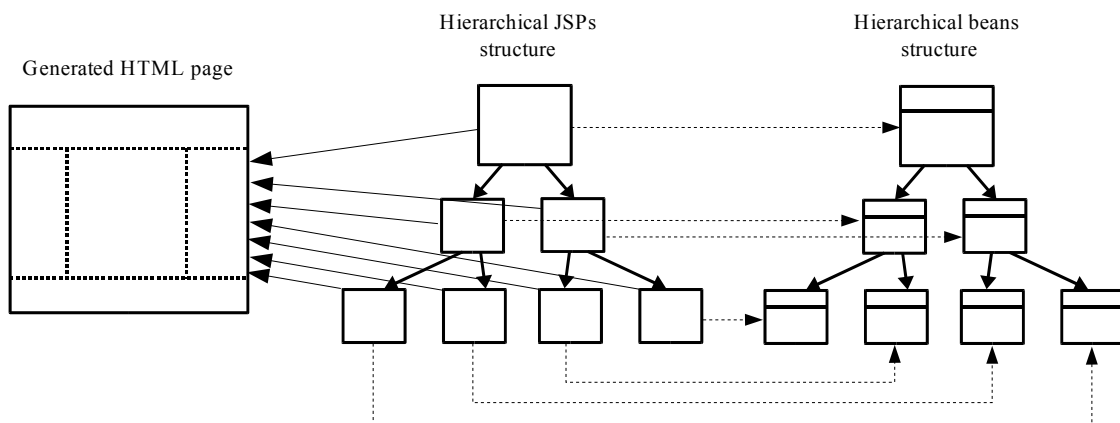


Figure 3. Hierarchical MVC architecture

---

## Related Works

---

Nowadays, there are several technologies that apply MVC architecture for developing web applications. Most representative in J2EE they are Struts [Struts, 2005] and Java Server Faces [JSF, 2005]. Both are very similar in the way they apply MVC architecture.

JSF applies it at page level; it is to say there is one JSP file per page. The model is represented by one object per page as well. The controller is implemented by the own technology and it is configured through defined tags in the JSP file. The controller updates JavaBeans properties with the values from, found in the interface components generated by the JSP file.

Struts uses JavaBeans in the model also and it implements the view with JSP files that generate the output from those JavaBeans. The controller manages the requests received by the web application. It associates the actions to logic names used to build the links and forms in the JSP file.

Both technologies apply navigational paradigm and manage links through configuration files.

Main difference between these technologies and our proposal is that they have the navigation and page concepts as their base. In our approach, the page concept is not applied; instead of, we build a view from the information in the web application. Interactions with user make not him to navigate but to change the state of the application and when it is presented again, user can see the changes. Due to this approach, there is not one JSP file per page and when an object needs to be presented then, it will use its own JSP file. If the state of the object is based in other objects, it can delegate the presentation to them including other JSP files.

---

## Conclusion

---

In high interactive web applications, it is not possible to apply in a natural way the navigational paradigm because the user needs to see the state of the application on every moment. The architecture proposed here presents a set of classes that shows their content and modifications at a given time, so not navigational point of view is used in this approach. This fact helps a lot to increase the manageability of the project at developing time. On the other hand, this approach takes good care of encapsulation and reusability benefits offered by classes saving time in the develop process.

In the future, incorporating this architecture into an Integrated Development Environment will enable faster developing giving to the programmer the possibility of generating big parts of his code automatically. Moreover, building the user interface hierarchically, so it can be used the same delegation as in object oriented technology, makes that repercussions in changing code of the application are very limited.

Next to this work, we will continue developing tools, code generators and libraries while going deeper into these concepts to form a complete framework that will help in development of web applications with a high interaction.

---

## Acknowledgement

---

We would like to thank the following people for their help and their support: Luis Fernández Muñoz, Fernando Arroyo Montoro, José Ernesto Jiménez Merino and Carmen Luengo Velasco.

---

## Bibliography

---

- [J2EE, 2005] Java 2 Enterprise Edition. <http://java.sun.com/j2ee/>
- [Krasner, 1988] Krasner, G. and Pope, S., 1988. A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 system. *Journal of Object Oriented Programming*, Vol. 1, No. 3, pp 26-49.
- [Swing, 2005] Swing. <http://java.sun.com/products/jfc/>
- [JavaBeans, 2005] JavaBeans. <http://java.sun.com/products/javabeans/>
- [JSP, 2005] JavaServer Pages. <http://java.sun.com/j2ee/jsp/>
- [Taglibs, 2005] Tag libraries. <http://java.sun.com/products/jsp/taglibraries/>
- [Struts, 2005] Struts. <http://jakarta.apache.org/struts/>
- [JSF, 2005] JavaServer Faces. <http://java.sun.com/j2ee/javaxserverfaces/>

---

## Authors' Information

---

**Micael Gallego Carrillo** – ESCET, Universidad Rey Juan Carlos, C/Tulipán s/n, 28933 – Móstoles (Madrid), Spain; e-mail: [mgallego@escet.urjc.es](mailto:mgallego@escet.urjc.es)

**Iván García Alcaide** – LPSI, Universidad Politécnica de Madrid; Campus Sur, Carretera de Valencia Km. 7, 28031 Madrid, Spain; e-mail: [igarcia@eui.upm.es](mailto:igarcia@eui.upm.es)

**Soto Montalvo Herranz** – ESCET, Universidad Rey Juan Carlos, C/Tulipán s/n, 28933 – Móstoles (Madrid), Spain; e-mail: [soto.montalvo@urjc.es](mailto:soto.montalvo@urjc.es)

## A SENSITIVE METRIC OF CLASS COHESION

Luis Fernández, Rosalía Peña

**Abstract:** *Metrics estimate the quality of different aspects of software. In particular, cohesion indicates how well the parts of a system hold together. A metric to evaluate class cohesion is important in object-oriented programming because it gives an indication of a good design of classes.*

*There are several proposals of metrics for class cohesion but they have several problems (for instance, low discrimination). In this paper, a new metric to evaluate class cohesion is proposed, called SCOM, which has several relevant features. It has an intuitive and analytical formulation, what is necessary to apply it to large-size software systems. It is normalized to produce values in the range [0..1], thus yielding meaningful values. It is also more sensitive than those previously reported in the literature. The attributes and methods used to evaluate SCOM are unambiguously stated. SCOM has an analytical threshold, which is a very useful but rare feature in software metrics. We assess the metric with several sample cases, showing that it gives more sensitive values than other well know cohesion metrics.*

**Keywords:** *Object-Oriented Programming, Metrics/Measurement, Quality analysis and Evaluation.*

---

## Introduction

---

The capacity to measure a process facilitates its improvement. Software metrics have become essential in software engineering for quality assessment and improvement. Class cohesion is a measure of consistency in the functionality of object-oriented programs. High cohesion implies separation of responsibilities, components' independence and less complexity. Therefore, it augments understandability, effectiveness and adaptability. Actually, these are major factors of the great interest in using object-oriented programming in software engineering.

“Current cohesion metrics are too coarse criteria that should be complemented with finer-grained factors (...). Then it will be easier to assess the trade off involved in any design activity, which would make it possible to see whether a system is evolving in the right direction [Mens, 2002].

One software metric must be simple, precise, general and computable in order to be applicable to large-size software systems. Automated metric producing sensitive values of class cohesion can be of great value to most designers, developers, managers, and of course to beginners, in identifying the class cohesion in an object-oriented design.

This paper presents, as follows, a metric of class cohesion that has several advantages over previous cohesion metric proposals. The next section presents a brief summary of the state of art about proposed metrics of class cohesion. In the third section, we formulate the new metric. The existence of thresholds is studied in fourth section, where two representative values are analytically determined. This section also provides guidelines to improve a class definition when its cohesion value is below the threshold. In fifth section, the method and

attributes to be considered on the evaluation of a class are discussed. The sixth section assesses qualitative and quantitatively the desirable features of our cohesion metrics. Particularly interesting is a comparative analysis with previous cohesion metrics, based on a sample of classes with different cohesion features. Finally, we summarize our conclusions.

---

## Related Work

---

In a cohesive class, all the methods collaborate to provide the class services and to manage the object state; in others words methods work intensively with the attributes of the class. A metric for class cohesion was first proposed by Chidamber and Kemerer in 1991 and then revised in 1994 [Chidamber, 1994]. Then, it has been repeatedly reinterpreted and improved [Li, 1993][Li, 1995][Hitz, 1996]. Properly speaking, these proposals do not evaluate cohesion but lack of cohesion (hence its name, LCOM) in terms of the number of pairs of the class methods that use disjoint sets of attributes.

There are two problems with Chidamber and Kemerer's expression [Chidamber, 1994]:

- i. the metric has no upper limit, so it is not easy to grasp the meaning of the computed value, and
- ii. there are a large number of dissimilar examples, all of them giving the same value (LCOM=0) but not necessarily indicating the same cohesion.

In spite of its pitfalls, LCOM still is the most widely used metric for cohesion evaluation [Bansiya, 1999]. Li et al. [Li, 1995] estimate the number of non-cohesive classes that should be redesigned by computing the number of subsets of methods that use disjoint subsets of attributes. We call this value the number of clusters of the class. Etzkorn et al. [Etzkorn, 1998] study which formulation provides the best interpretation of lack of class cohesion.

Henderson-Sellers [Henderson-Sellers, 1996] proposed a completely new expression for cohesion evaluation, called LCOM\*:

$$LCOM^* = \frac{\left[ \frac{1}{a} \sum_{j=1}^a \mu(A_j) \right] - m}{1 - m} \quad (1)$$

where  $a$  and  $m$  are the number of attributes and methods of the class, respectively, and  $\mu(A_j)$  is the number of methods that access the datum  $A_j$  ( $1 \leq j \leq a$ ). Notice that this expression is normalized to the range  $[0..1]$ . He claims that this metric has the advantage of being easier to compute than previous ones.

A problem with LCOM\* is that it is not clear how it can give any account of class cohesion or inter-method cohesion (relation between methods). In effect, the summatory in the numerator just counts how many times all the attributes are accessed, independently of which method accesses each datum. This problem is illustrated in Table 1, with a comparison of several metrics, including ours. The table includes a representation of classes in the second column. As well, classes are represented by another table in which rows can be found the attributes and in the columns the methods accessed by them. So, a filled cell $_{i,j}$  means that the  $i$ -th method uses the  $j$ -th attribute. For example, class C has 6 attributes and 6 methods, and its method  $M_4$  uses attributes  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$ .

In particular, rows F, G and H show that whereas class F seems to be a good candidate to be split into two classes, it is not clear whether a more cohesive design can be found for classes G or H. However, the three classes are evaluated by LCOM\* with the same value, it seems to be clear that it does not give a good account of class cohesion at least with these two examples.









Metrics measuring lack of cohesion, as proposed in [Chidamber, 1994][Li, 1993][Li, 1995][Hitz, 1996], evaluate the number of disjoint pairs of methods. From the opposite point of view, cohesion is related to the overlapping of these sets. It would be desirable to measure not only whether two methods are disjoint, but also, how big the intersection set is. Let us consider that two methods are more cohesive if they co-use more attributes. Consequently, cohesion is directly proportional to the cardinality of intersection for all the possible pairs of methods. [Park, 1998] suggested an evaluation of Class Cohesion (CC) as the sum of the connection intensity

between all the possible pairs of methods. Let us denote  $I_k$  for the subset of attributes used by a given method  $M_k$ . The connection intensity of a pair of methods is defined as:

$$C_{i,j} = \frac{\text{Card}(I_i \cap I_j)}{a} \quad (2)$$

where  $a$  is the total number of instance attributes in the class. CC is more intuitive than LCOM\*, but it also has deficiencies, as is shown in Table 1, where it also computes the same value for F, G, H classes.

Table 1. Values of different cohesion metrics on 8 sample classes.

CLASS		LACK of COHESION			COHESION	
Name	Methods/Attribute Table	Chidamber (1994)	Li (1995)	Henderson (1996)	Park (1998)	SCOM
A		15	6	1	0	0
B		0	1	0	1	1
C		0	1	0.23	0.62	0.82
D		0	1	0.57	0.23	0.29
E		3	2	0.60	0.20	0.20
F		2	2	0.67	0.17	0.17
G		0	1	0.67	0.17	0.25
H		0	1	0.67	0.17	0.36

### Sensitive Class Cohesion Metric (SCOM)

In this section, we introduce a new cohesion metric that we call SCOM (standing for Sensitive Class Cohesion Metric). At this stage, let  $\{A_1, \dots, A_a\}$  and  $\{M_1, \dots, M_m\}$  be the set of attributes and methods of the class, respectively, to be considered.

Firstly, let us consider connection intensity of a pair of methods (see formula 2 above). It would be more accurate if it were divided by the maximum possible value, rather than by the total number of attributes of the class, as proposed by Park in formula 2. As the largest possible cardinality of the intersection of two sets is the minimum of their cardinalities, the connection intensity formula becomes:

$$C_{i,j} = \begin{cases} 0, & \text{if } I_i \cap I_j = \phi \\ \frac{\text{card}(I_i \cap I_j)}{\min(\text{card}(I_i), \text{card}(I_j))} & \text{otherwise} \end{cases}$$

Connection intensity of a pair of methods must be given more weight when such a pair involves more attributes. Of course, the largest contribution is found when every attribute in the class is used by any of the two methods. The weight factor is:

$$\alpha_{i,j} = \frac{\text{card}(I_i \cup I_j)}{a}$$

Finally, the formula of our cohesion metric results:

$$SCOM = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m C_{i,j} * \alpha_{i,j} \quad (3)$$

where the coefficient of the summatory is the inverse of the total number of method pairs:  $\binom{m}{2} = \frac{m(m-1)}{2}$ . As

a consequence, the metric is normalized to the range [0..1], where the two extreme values are:

- i. Value zero: there is no cohesion at all, i.e. every method deals with an independent set of attributes.
- ii. Value one: full cohesion, i.e. every method uses all the attributes of the class.

## Thresholds

The threshold or “alarm value” minimum (respectively, maximum) of a metric is a value that indicates a design problem for a software entity whose evaluation lies below (or over) it. The threshold calls for the developer’s attention to focus on a particular module or chunk for further evaluation [Lorenz, 1994].

As we mentioned in the previous section, the SCOM cohesion metric is ranged between zero (representing no cohesion at all) and one (representing full cohesion). Although an empirical threshold must be determined with real projects, we consider that a theoretical study about singular points gives information about how to interpret the values the metric delivers. In the next two subsections, we present the analytical expressions for the minimal value of a class that has one cluster and the maximum value of a class that has at least two clusters. The third subsection deals with the influence of these values on the SCOM threshold and its applicability for class evaluation.

### Minimal cohesion value for one cluster.

Let us consider a class with  $m$  methods and  $a$  attributes defined so that it is impossible to find two clusters.

We start with  $m=2$  and then we include one by one new methods to the class satisfying the former condition. The lower contribution to cohesion is given by a method with just one attribute. The connection intensity  $C_{ij}$  of a pair of methods with only one attribute has the denominator equal to one (see formula 3). Counting the number of non-null terms, we induce the total number of pairs of the class with  $m$  methods and  $a$  attributes.

Table 2 shows the minimum number of pairs of non-null methods for the values of  $a=3$  and  $a=4$  and consecutive values of  $m$  starting from  $m=2$ . These minimums are expressed in terms of  $a$ ’s and  $m$ ’s. From this table, the general expression establishing the number of terms that contributes to cohesion in a significant manner can be deduced by induction.

Table 2. Number of pairs non-null methods.

m	a=3		a=4	
	S	S(a,m)	S	S(a,m)
2	1	a·0+1	1	a(0)+1
3	2	a·0+2	2	a(0)+2

4	3	$a \cdot (0+1)$	3	$a(0)+3$
5	5	$a(0+1)+2$	4	$a(0+1)$
6	7	$a(0+1)+4$	6	$a(0+1)+2$
7	9	$a(0+1+2)$	8	$a(0+1)+4$
8	12	$a(0+1+2)+3$	10	$a(0+1)+6$
9	15	$a(0+1+2)+6$	12	$a(0+1+2)$
10	18	$a(0+1+2+3)$	15	$a(0+1+2)+3$
...	...	...	...	...

From these data, it can be induced that the number of terms that contributes in a significant manner to cohesion, named  $S$ , has a linear behavior. Hence,  $S$  can be expressed by  $S = a \cdot t + r$ , where:

$$t = \frac{1}{2} \text{INT}\left(\frac{m-1}{a}\right) \cdot \left[ \text{INT}\left(\frac{m-1}{a}\right) + 1 \right]$$

and

$$r = \left[ 1 + \text{INT}\left(\frac{m-1}{a}\right) \right] \cdot \text{Mod}\left(\frac{m-1}{a}\right)$$

hence,

$$S = \frac{a}{2} \text{INT}\left(\frac{m-1}{a}\right) \cdot \left[ \text{INT}\left(\frac{m-1}{a}\right) + 1 \right] + \left[ 1 + \text{INT}\left(\frac{m-1}{a}\right) \right] \cdot \text{Mod}\left(\frac{m-1}{a}\right)$$

and extracting common factors:

$$S(m, a) = \frac{1}{2} \left[ 1 + \text{int}\left(\frac{m-1}{a}\right) \right] \left[ \text{mod}\left(\frac{m-1}{a}\right) + m - 1 \right]$$

The minimum weight factor  $a_{i,j}$  occurs when both methods deal with the same attribute being  $1/a$ , but some coefficients will actually be higher. Therefore:

$$SCOM \min > \frac{2}{m(m-1)a} S(m, a) = SCOM \min K \quad (4)$$

We call  $SCOM_{\min K}$  to the minimal known value. A class with  $SCOM < SCOM_{\min K}$  does not satisfy the induction premise. Consequently, we can claim that it has at least two clusters and it must be subdivided into smaller, more cohesive classes.

#### Maximum cohesion value for two clusters.

From the opposite point of view, it is possible to evaluate the maximum value of  $SCOM$  in presence of two clusters. In this situation there are at least  $m-1$  null terms in (3); in other words, there are at most  $(m-1) \cdot (m-2)/2$  non-null terms. The biggest  $C_{i,j}$  is valued one. The biggest weight factor occurs when one cluster has  $a-1$  attributes and the other has just one attribute and all the methods in each cluster involve all the attributes in such a cluster. In this situation,  $a_{i,j} = (a-1)/a$ .

Finally, the maximum value of  $SCOM$  for a class with  $m$  methods arranged in two clusters is:

$$SCOM 2 \max = \frac{(a-1)(m-2)}{ma} \quad (5)$$

#### Threshold applicability.

As we explained above, the equation (4) provides the minimum value below which certainly there are two subsets using non-overlapping attributes. This class can be automatically split into two (or more) smaller and simpler classes.

The metrics for class size by Lorenz and Kidd [Lorenz, 1994] suggest a threshold of 20 instance methods and 4 class methods, being 12 the number average of methods per class in a project. They also proposed 3 as a threshold both for instance attributes and class attributes in a class, and 0, 1 the class attributes average in a project. With these particular values ( $a=3$  and  $m=12$ ), the minimal cohesion is  $SCOM_{\min K}=0.13$ .

Moreover,  $SCOM_{2\max}$  provides the analytical value that guarantees the existence of one cluster. For  $a=3$  and  $m=12$ ,  $SCOM_{2\max}=0.56$ . Figure 1 places both values on the cohesion range.

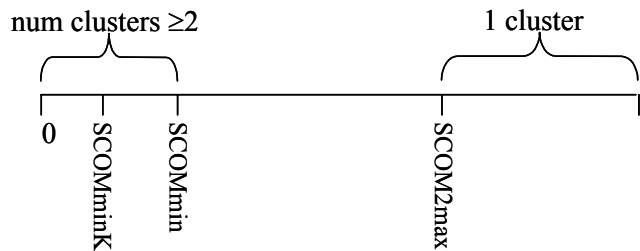


Figure 1. Significant values on cohesion range.

The presence of two clusters clearly indicates that the class may be split. Not all the classes with more than one cluster can be detected by  $SCOM_{\min K}$ . Li formulation could help but, even in presence of just one cluster, the class design can be low cohesive and SCOM is giving account of such a situation.

The threshold considered in software evaluation should be between  $SCOM_{\min K}$  and 1. It looks feasible to demand guarantee of just one cluster, then SCOM will actually be between  $SCOM_{2\max}$  and 1. The adequate value, inside these limits, should be established by studying real projects.

Quoting Lorenz and Kidd [Lorenz], "metrics are guidelines of the quality of the design and not absolute laws of nature". Thus, when the alarm is raised, the corresponding software entity is suspicious to have design problems, so it must be re-analyzed. However, thresholds are affected by many factors (as pointed out in the next section), so the warning may sometimes be disregarded.

## Criteria of Application

The features of object-oriented programming do not always make obvious what methods and attributes to consider for estimating cohesion. Moreover, different authors use different criteria and even they say nothing about what to do with them. This un-definition or not precise definition in the way of evaluating cohesion makes very difficult to interpret the obtained values from the metrics. We discuss here the most relevant aspects.

- i. Transitive closure of attributes. Whenever  $M_i$  invokes method  $M_j$ ,  $M_i$  transitively uses each attribute being used by  $M_j$ . Consequently, these attributes must be taken into account for the metric evaluation. The same reason is argued to consider objects that are exported to clients which map a subsequent client request (method invocation sending the argument "this"), as it happens in the double dispatching and visitor patterns. Attributes involved in subsequent request must be counted for the exporter method.
- ii. Class and instance methods and attributes. Class cohesion involves both instance and class methods and attributes. It is common that class attributes are less involved in methods than instance attributes, which would decrease inadequately the class cohesion. These attributes could be removed to evaluate cohesion of a single class. However, the rate between class and instance attributes use to be too small in the average to affect evaluation substantially, so they can either be removed or considered.
- iii. Constructor and destructor methods. Etzkorn *et al.* [Etzkorn, 1998] suggested to exclude constructor methods but to keep the destructor method for LCOM calculations. The argument was that these particular functions often access most or all of the variables of a class. As LCOM looks for methods with



non-overlapping attributes, a method using all the attributes lowers its capability of discrimination. However, we think that both constructor and destructor methods must be considered. For instance, a constructor that initializes all the attributes (as probably happens in the Car class of a concessionaire) contributes to a larger cohesion value than a constructor that just initializes some attributes (as will likely happen in a Student class, where the name will be required from the very beginning, but the student final evaluation will not).

- iv. Access methods. The *get* and *put* private methods make easier future maintenance. They typically cope with just one attribute. However, the resulting cohesion value should not be negatively affected by a good programming style. Therefore, we suggest excluding them.
- v. Methods overload. It is common in object-oriented programming to have a given method called with a concrete value. Some programming languages (such as C++) allow parameter initialization in absence (for example, an interval class being displaced to the coordinate origin). When a language lacks this feature (for instance, in Java or Smalltalk), the method is overloaded by defining a new method without this parameter declaration, where the new method's implementation just consists in the invocation of the former with a constant value. Both methods cope with the same functionality and in this particular case, only the general one must be considered for cohesion evaluation.
- vi. Inheritance in cohesion evaluation. The question of whether inherited variables must be considered for cohesion evaluation was posed by Li *et al.* [Li, 1995]. They only considered local attributes. However, in [Etzkorn, 1998] argues in favor of the consideration of inherited attributes. The state of one instance is defined by its own and parent attributes, so both of them must be considered. Thus, we propose to consider the parent's methods and attributes in the parent class cohesion evaluation, and the parent's and its own attributes, but just its own methods being involved on subclass cohesion evaluation. In some types of inheritance, the subclass methods seldom use the inherited attributes; in such a situation, the metric value decreases.
- vii. Methods dealing with no attributes. We find this situation in abstract classes and sometimes in static methods. On the one hand, abstract methods must not be considered for cohesion evaluation of the abstract class. They must be considered for the subclass where they are actually implemented. On the other hand, static methods not dealing with attributes (as happens in the factory pattern) are not considered for cohesion. It is expected that instance methods manage at least one attribute; otherwise, a different metric from cohesion must be used and it must raise the appropriate alarm. Therefore, methods, which deal with no attributes, must not be considered.
- viii. Attributes not used by any method. This is an undesirable situation in absence of inheritance. Actually, some compilers (e.g. Fortran90 or Java) warn about it. A different metric will deal with it. In presence of inheritance, the parent class sometimes references an attribute to be managed by its subclasses. These attributes must not be considered for superclass evaluation, but they must be considered for subclass evaluation.

---

## Metric Assessment

---

In this section, we assess our SCOM metric. Firstly, we make an informal assessment. Then, we make an empirical comparison with other metrics.

### Qualitative analysis of cohesion metrics.

The desirable properties (or metametrics) of a software metric have been proposed elsewhere [Henderson-Seller, 1996][Xenos, 2000]. We assess SCOM informally with respect to such properties:

- i. Simplicity (the definition of the metric is easily understood). SCOM has a simple definition.
- ii. Independence (the metric does not depend on its interpretation). SCOM is independent because it has a clear definition based on syntactic information (attributes and methods).
- iii. Accuracy (the metric is related to the characteristic to be measured). SCOM is accurate since it is based on cohesion features, namely dependencies between attributes and methods.

- iv. Automation (effort required to automate the metric). SCOM has an analytical definition that depends on the attributes used by the methods in a class and their transitive closure. This information can be automatically extracted from the class implementation.
- v. Value of implementation (the metric may be measured in early stages or it evaluates the success of the implementation). SCOM measures the class implementation.
- vi. Sensitivity (the metric provides different values for non-equivalent elements). SCOM is sensitive for different cases, as it is shown in the next subsection.
- vii. Monotonicity (the value computed for a unity is equal or greater/less than the sum of the values of its components). SCOM is decreasingly monotonous.
- viii. Measurement scale (rate/absolute/nominal/ordinal, being more useful rate and absolute scales than nominal and ordinal ones). SCOM gives a normalized rate scale.
- ix. Boundary marks (upper and/or lower bounds). SCOM has bounds on both sides of the scale since it ranges between 0 and 1.
- x. Prescription (the metric describes how to improve the entity measured). SCOM establishes how to improve the class as we describe below.

There are two typical cases where some restructuring actions can be performed on a class. Firstly, when two (or more) non-overlapping clusters are detected in a class, it must be split into two (or more) classes. The methods using a subset of non-overlapping attributes must be grouped, taking care to preserve the appropriate granularity (size) of the class.

Secondly, when the threshold alarm is raised but isolated clusters are not detected, the class is a candidate to be re-analyzed. The programmer must inquire whether any attribute can be moved to another class, look for any undiscovered class, or check whether all the class responsibilities are necessary (or any can be moved to a different class). By doing this clean up, some methods may go away or may be shortened. As a consequence, the comprehensibility of the application and the quality of the class design are improved, while code length actually decreases.

#### **Quantitative analysis of cohesion metrics.**

We have applied different cohesion metrics, including SCOM, to several sample classes in order to empirically comparing their results. Table 1 illustrated such results for eight sample classes.

Notice that while Chidamber, Li-Henry and Henderson evaluate lack of cohesion, Park and SCOM evaluate cohesion, so their extreme values have the opposite interpretation.

Class A has no cohesion at all while class B has the highest cohesion, so they obtain the limit values for metrics with bounds (Henderson, Park and SCOM).

Classes B, C and D represent very different situations. Chidamber and Li-Henry metrics show low sensibility, rating the same value (zero) for these classes, whereas the other metrics give an account of their differences.

Classes E and F are very similar: they share the property of having two clusters, being obvious that they could be split into two. Chidamber's metric yields different values for these classes, showing that the meaning of non-zero values in this formulation are difficult to understand because it does not have an upper bound. Henderson's metric also exhibits low discrimination because the lack of cohesion measure obtained is not as high as we could expect.

Classes F, G and H also represent different situations. Henderson and Park cohesion metrics yield the same value, showing lower discrimination capability than expected. However, SCOM gives account of their differences, yielding the lowest value for the obviously worst design (F).

SCOM is more sensitive than Park's formulation, as can be seen comparing the cases D and E. Moreover, Park's formulation seems to be too restrictive for the cohesion's upper extreme, as it is shown in class C. This class has a good cohesive appearance, so a value higher than 0.62 is intuitively expected for a magnitude ranging between 0 and 1. SCOM behaves better, yielding the value 0.82.

Remember that our metric also allows computing the analytical values  $SCOM_{minK}$  and  $SCOM_{2max}$  that give information about the cohesion for a particular class, without any previous human analysis of its table structure. For a class with 4 attributes and 4 methods,  $SCOM_{minK}=0.13$  and  $SCOM_{2max}=0.38$ . It can be seen that classes F,

---

G and H are low cohesive; in particular, F should be split. For a class with 6 attributes and 6 methods,  $SCOM_{2max}=0.56$ . The classes D and E have values far below  $SCOM_{2max}$  (0.29 and 0.20, respectively), indicating design problems. The value 0.82 obtained for class C suggests a good cohesive design, as expected analyzing its methods/attributes table.

---

## Conclusion

This paper proposes a metric to evaluate class cohesion, that we have called SCOM, having has several relevant features. It has an intuitive and analytical formulation. It is normalized to produce values in the range [0..1]. It is also more sensitive than those previously reported in the literature. It has an analytical threshold, which is a very useful but rare feature in software metrics. The attributes and methods used for SCOM computation are also unambiguously stated.

The metric is simple, precise, general and amenable to be automated, which are important properties to be applicable to large-size software systems. By following the prescription provided by the metric, the understanding and design quality of the application can be improved.

---

## Bibliography

- [Mens, 2002] T.Mens and S.Demeyer. Future trends in software evolution metrics. In: Proc. IWPSE2001, ACM, 2002, pp. 83-86.
- [Chidamber, 1994] S.R.Chidamber, and C.F.Kemerer. A Metrics Suite for Object Oriented Design. In: IEEE Transactions on Software Engineering, 1994, 20(6), pp. 476-493.
- [Li, 1993] W.Li, and S.M.Henry. Maintenance metrics for the object oriented paradigm. In: Proc. 1st International Software Metrics Symposium, Baltimore, MD: IEEE Computer Society, 1993. pp. 52-60.
- [Li, 1995] W.Li, S.M.Henry et al. Measuring object-oriented design. In: Journal of Object-Oriented Programming, July/August, 1995, pp.48-55.
- [Hitz, 1996] M.Hitz, and B.Montazeri. Chidamber & Kemerer's metrics suite: a measurement theory perspective. In: IEEE Transactions on Software Engineering, 1996, 22(4), pp. 267-271.
- [Bansiya, 1999] J.Bansiya, L.Etz Korn, C.Davis, and W.Li. A Class Cohesion Metric For Object-Oriented Designs. In: JOOP, 1999, 11(8), pp. 47-52.
- [Etz Korn, 1998] L.Etz Korn, C.Davis and W.Li. A practical look at the lack of cohesion in methods metric. In: JOOP, 1998, 11(5), pp. 27-34.
- [Henderson-Sellers, 1996] B.Henderson-Sellers. Object-Oriented Metrics: Measures of Complexity. In: New Jersey, Prentice-Hall, 1996, pp. 142-147.
- [Park, 1998] S.Park, E.S.Hong et al. Metrics measuring cohesion and coupling in object-oriented programs. In: Journal of KISS, 1998, 25(12), pp. 1779-87.
- [Lorenz, 1994] M.Lorenz, and J.Kidd. Object-Oriented Software Metrics: A Practical Guide. In: New Jersey, Prentice Hall, 1994, p. 73.
- [Xenos, 2000] M.Xenos, D.Stavrinoudis, K.Zikouli, and D.Christodoulakis. Object-Oriented Metrics: A Survey. In: Proc. FESMA, Madrid, 2000, pp. 1-10.

---

## Authors' Information

**Luis Fernández** – UPM, Universidad Politécnica de Madrid; Ctra Valencia km 7, Madrid-28071, España; e-mail: [setillo@eui.upm.es](mailto:setillo@eui.upm.es)

**Rosalía Peña** – UA, Universidad de Alcalá; Ctra Madrid/Barcelona km 33, Alcalá-28871, España; e-mail: [rpr@uah.es](mailto:rpr@uah.es)

## RKHS-METHODS AT SERIES SUMMATION FOR SOFTWARE IMPLEMENTATION

**Svetlana Chumachenko, Ludmila Kirichenko**

**Abstract:** *Reproducing Kernel Hilbert Space (RKHS) and Reproducing Transformation Methods for Series Summation that allow analytically obtaining alternative representations for series in the finite form are developed.*

**Keywords:** *The reproducing transformation method, Hilbert space, reproducing kernel, RKHS, Series Summation Method.*

---

### Introduction

---

Operating speed of digital logic devices depends on type of silicon: PLD, Gate Array or ASIC. FPGAs are the lowest in risk, low in project budget but have the highest cost per unit. Gate Arrays utilize less custom mask making than standard cell and stand in the middle from all of three and fallen from wide use today. Cell based ASICs have the highest performance and lowest cost per unit in case of mass production, but they also have the longest and most expensive design cycle. Also, digital designs can be divided on CPU based systems on chip (SoC) and non-CPU logic devices. CPU as universal processing unit can solve broad spectrum of various tasks from all areas of human activity. Nevertheless, there exist bottlenecks where CPU can't satisfy required performance. Usually it happens during implementation of mathematical tasks that require big number of iterations and hence big time expenses to obtain desired result with desired accuracy.

To increase efficiency of solving of computational tasks there are used mathematical co-processors, which implement most efficient ways of computing equations, integrals, differential coefficients. It is obvious that after discovering of new methods of increasing computation accuracy and decreasing computation time it is necessary to re-implement mathematical co-processors or use new generation of IP-cores in PLD, Gate Array, ASIC designs. It is presented, easy to implement as IP-core, method of reduction of computation of certain types of series to exact function, that is widely used during calculation of parameters of high radio frequency devices. Presented method decrease computation time of such tasks in tens and hundred times and its inaccuracy is equals to zero.

---

### Statement of the Problem

---

The investigation is based on fundamental works of Aronzajn [1], Razmahnin, Yakovlev [2]. It develops the following research [10-12] on Series Summation in Reproducing Hilbert Space and their approbation [19-22]. Modern papers of Saitoh, Laslo Mate, Daubeshies and others [3-9] are used for revues and staying problem. Classical papers of Tranter, Doetsch, Ango, Titchmarch, Ahiezer [13-18] are used for inter-comparison of results.

Mathematical models based on Reproducing Kernel Hilbert Space methods are used in Wavelets Analysis, namely: at Pattern Recognition, Digital Data Processing, Image Compression, Computer Graphics; and also in Learning theory: for example, at Exact Incremental Learning, in Statistical Learning theory, in Regularization theory and Support Vector Machines. In mentioned arias we have not deal with exact Series Summation because it isn't necessary for considered cases. We use sum and finite summation, not series. But there are areas of scientific study where exact series summation it is necessary.

For such problems Reproducing Transformations Method and its part – Series Summation Method in RKHS – can be useful [20, 22]. We are going to point out these areas.

*The purpose* of the investigation is to originate a new Series Summation Method based on RKHS-theory and to demonstrate the new results which develop theoretical statements of Series Summation Method in RKHS.

*The research problems* are:

- Series Summation in RKHS
- Applications of Series Summation in RKHS
- Reproducing Transformations Method as a perspective of this research

### Base Theoretical Statements and Investigation Essence

Reproducing Kernel Hilbert Space is a subspace of Hilbert space with Reproducing Kernel (RK). RK is a function Ker of two variables with two properties: 1)  $\forall t \in T \text{ Ker}(s_0, t) \in T$ ; 2)  $\forall f \in H \text{ } f(t) = \langle f(s), \text{Ker}(s, t) \rangle$ , where  $\langle \dots \rangle$  – inner (scalar) product can be represented as a series on selective values. There is an operator G, which transfers any function from Hilbert space  $L_2$  into function from RKHS and leaves without change function from RKHS H.

Thus, there is an operator G, which transfers any function from Hilbert space  $L_2$  into function from RKHS and doesn't change any function from RKHS H.

For example, the functions with finite spectrum of cosine- and sine-transformations and Hankel-transformation form RKHS. The basic research of expansion problem on selective values was executed by K. Shannon and V.A. Kotelnikov. There are statements determining particular cases RKHS. Thus, any function from RKHS can be represented as selective value expansion. If there is series where the common summand can be reduced to a standard form, – it means to extract reproducing kernel by equivalent transformations, – then for any series one can put in accordance a function from RKHS. In other words, a series can be summarized by known formulas.

Thus, the main idea of proposed method is to obtain and to use the following relation:

$$f(s) = \sum_k f(t_k) \text{Ker}(s, t_k),$$

in right-hand side of this relation we can see a series on selective values of function f(t); left-hand side represents value of function f in point s.

We use four base kinds of Reproducing Kernels, which originate four RKHS accordingly [21]:

1. RKHS  $H_1$  is a space of functions, which have finite Fourier-transformations.
2. Space  $H_2$  contains a class of functions with finite Hankel-transformation.
3. Space  $H_3$  consists of functions with finite sine-transformations.
4. Space  $H_4$  has all functions, those cosine-transformations are finite.

For these spaces there are four Kernel Functions and Series on selective values accordingly [21].

Based RKHS-theory the new approach to definition of series sum is proposed. It is called *Series Summation Method in RKHS*.

It allows analytically obtaining alternative representations for some kinds of series in the finite form [10].

The new formulas for calculating the sum of series (including alternating) have been obtained by proving several theorems [10, 12].

*Reproducing Transformations Method* are generalization and extension of Series Summation method in RKHS [20, 22]. It can be useful at solving mentioned problems and other important points. It needs further evolution and consideration.

For example, we can see proving the following formula.

**Theorem 1.** There is the following relation for alternating series

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^n - k^n} = \begin{cases} \frac{\pi a F(a)}{a - k}, & n = 1; \\ \frac{\pi F(a)}{2 p a^{2p-2} \sin \pi a}, & n = 2p, p = 1, 2, 3, \dots; \\ \frac{\pi F(a)}{(2p-1) a^p \sin \pi a}, & n = 2p-1, p = 2, 3, \dots \end{cases} \quad (1)$$

for any  $F(x)$  from RKHS,  $a \neq 0, \pm 1, \pm 2, \dots$

Proof. Let's consider alternating series the common member of which contains the difference of powers in denominator:

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^n - k^n}.$$

Let's define its sum. For this purpose we would consider the cases, when  $n$  is equal to natural number.

1) Let  $n = 1$ . The common member of series transforms to kind [10] that yields:

$$\begin{aligned} \sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a-k} &= \sum_{k=1}^{\infty} F(k) \frac{\cos(k\pi) \sin(\pi a)}{(a+k) \sin \pi a} \frac{a+k}{a-k} \frac{2\pi k}{2\pi} = \frac{\pi}{2 \sin \pi a} \sum_{k=1}^{\infty} \frac{2k}{a+k} \frac{\sin \pi(a-k)}{\pi(a-k)} (a+k)F(k) = \\ &= \frac{\pi}{2 \sin \pi a} [ (a+k)F(k) ] \Big|_{k=a} = \frac{2\pi a F(a)}{2 \sin \pi a} = \frac{\pi a F(a)}{\sin \pi a}. \end{aligned}$$

2) For  $n = 2$  the result obtained in [10]:

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^2 - k^2} = \frac{\pi}{2 \sin \pi a} F(a). \quad (2)$$

3) For  $n = 3$  we can obtain result by recurrent way with accounting formula (2) and using the decomposition of difference of cubes:

$$\begin{aligned} \sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^3 - k^3} &= \sum_{k=1}^{\infty} (-1)^k k \frac{F(k)}{(a-k)(a^2 + ak + k^2)} = \sum_{k=1}^{\infty} \frac{(-1)^k k}{(a^2 - k^2)(a^2 + ak + k^2)} (a+k)F(k) = \\ &= \left[ \Phi(k) = \frac{(a+k)F(k)}{(a^2 + ak + k^2)} \right] = \frac{\pi}{2 \sin \pi a} \Phi(a) = \frac{\pi}{2 \sin \pi a} \frac{2aF(a)}{3a^2} = \frac{\pi F(a)}{3a \sin \pi a}. \end{aligned}$$

4) For  $n = 4$  we can obtain:

$$\begin{aligned} \sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^4 - k^4} &= \sum_{k=1}^{\infty} \frac{(-1)^k k}{(a^2 - k^2)(a^2 + k^2)} F(k) = \left[ \Phi(k) = \frac{F(k)}{(a^2 + k^2)} \right] = \frac{\pi}{2 \sin \pi a} \Phi(a) = \\ &= \frac{\pi}{2 \sin \pi a} \frac{F(a)}{2a^2} = \frac{\pi F(a)}{4a^2 \sin \pi a}. \end{aligned}$$

5) For  $n = 5$  we can obtain with decomposition by difference of fifth powers the following result:

$$\begin{aligned} \sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^5 - k^5} &= \sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{(a-k)(a^4 + ka^3 + k^2a^2 + k^3a + k^4)} = \\ &= \sum_{k=1}^{\infty} \frac{(-1)^k k}{(a^2 - k^2)(a^4 + ka^3 + k^2a^2 + k^3a + k^4)} (a+k)F(k) = \\ &= \left[ \Phi(k) = \frac{(a+k)F(k)}{(a^4 + ka^3 + k^2a^2 + k^3a + k^4)} \right] = \frac{\pi}{2 \sin \pi a} \Phi(a) = \frac{\pi}{2 \sin \pi a} \frac{2aF(a)}{5a^4} = \frac{\pi F(a)}{5a^3 \sin \pi a}. \end{aligned}$$

6) For  $n = 6$  we can analogically obtain:

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^6 - k^6} = \frac{\pi F(a)}{6a^4 \sin \pi a}.$$

Thus, based on mentioned transformations we can conclude:

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a-k} = \frac{\pi a F(a)}{a-k}, \quad n=1;$$

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^{2p} - k^{2p}} = \frac{\pi F(a)}{2pa^{2p-2} \sin \pi a}, \quad n = 2p, p = 1, 2, 3, \dots;$$

$$\sum_{k=1}^{\infty} (-1)^k \frac{kF(k)}{a^{2p-1} - k^{2p-1}} = \frac{\pi F(a)}{(2p-1)a^p \sin \pi a}, \quad n = 2p-1, p = 2, 3, \dots$$

Thus, the theorem 1 has been proof.

The following examples illustrate application of the theorem 1.

**Example 1.** To proof of the following formula truth:

$$\sum_{k=1}^{\infty} (-1)^k \frac{k \sin kx}{a^2 - k^2} = \frac{\pi \sin ax}{2 \sin \pi a}, \quad -\pi < x < \pi, a > 0, a \neq 1, 2, \dots,$$

the residues theory is used in [23]. However, application of the theorem 1 gives the same result:

$$\sum_{k=1}^{\infty} (-1)^k \frac{k \sin kx}{a^2 - k^2} = \frac{\pi}{2 \sin \pi a} \sin kx \Big|_{k=a} = \frac{\pi \sin ax}{2 \sin \pi a}, \quad a > 0, a \neq 1, 2, \dots$$

**Example 2.** To proof of the following identity truth Laplace transformation is used [24]. However application of the theorem 1 reduces to the same result but it is simpler solution:

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1} k}{k^2 - a^2} J_{2n+1}(kx) = \sum_{k=1}^{\infty} \frac{(-1)^k k}{k^2 a^2 - k^2} J_{2n+1}(kx) = \frac{\pi}{2} J_{2n+1}(ax) \operatorname{cosec}(\pi a), \quad -\pi < x < \pi.$$

**Example 3.** To proof of the following identity truth

$$\sum_{k=1}^{\infty} (-1)^k \frac{k \cos kx}{a^4 - k^4} = \frac{\pi \cos ax}{4a^2 \sin \pi a}, \tag{3}$$

we can apply the theorem 1. Also we can show numerically this result (see fig. 1, 2). On Fig. 1 there are two diagrams in the equal co-ordinates for parameter  $a = 2,5$ . Graphs of function from right-hand side of (3) (the bold curve) and left-hand side (thin curve) of (3). Fig. 2 demonstrates the absolute uncertainty.

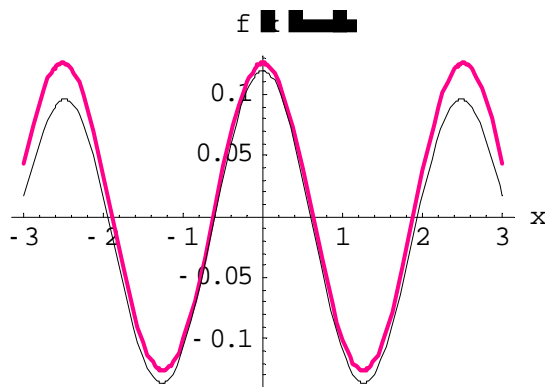


Fig. 1. Results comparison

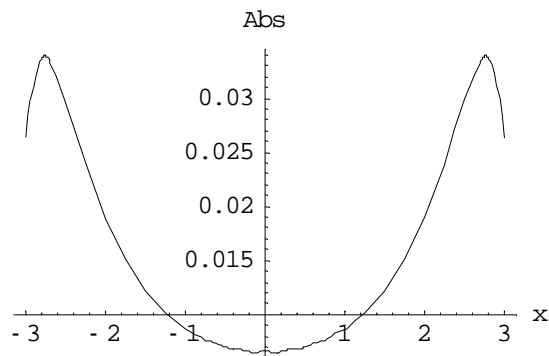


Fig. 2. Absolute error

---

## Conclusion

Thus, we can direct the following areas for applications of new Series Summation Method:

- Exact summation of series;
- solving summatory equations and its systems;

- solving integral equations and its systems;
- solving integral-summatory equations and systems of complex form;
- proving integral identities.

Mentioned areas can be used at solving some problems of: antenna theory; diffraction theory; electrodynamics and can be useful at Software/Hardware implementations (See Fig. 3).

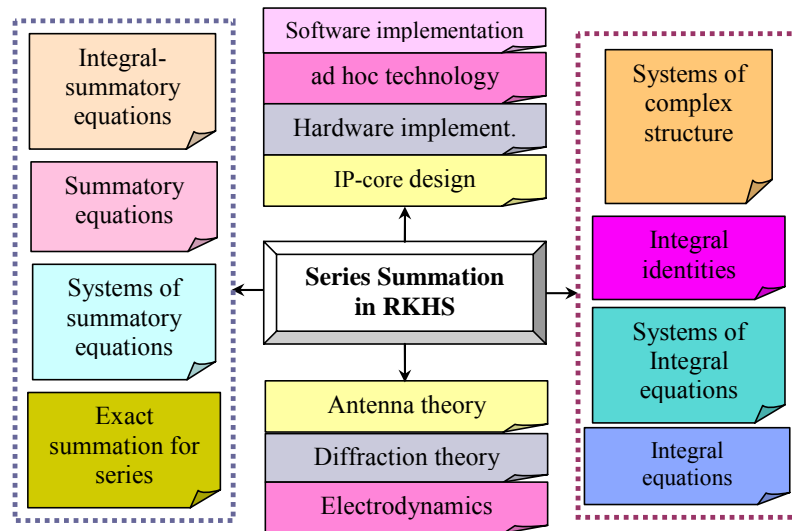


Fig. 3. Application arias of Series Summation in RKHS

The obtained results allow making the following *conclusions*:

- 1) RKHS-theory can be used for summation of selected series. For this purpose, *Series Summation Method in RKHS* has been proposed.
- 2) *Advantages of this method* consist of:
  - application of equivalent transformations to the common member of a series, that enables to obtain the *analytical solution* for smaller quantity of steps;
  - in absence of necessity to use the tables of integral transformations and to use the integration in complex area.
- 3) The application of obtained results of RKHS-theory for *solving the boundary electrodynamics problems* gives possibility to *simplify known methods* and to receive on their basis the *analytical solutions*, that is represented as essential for the further numerical experiment;
- 4) New mathematical results for solving the summatory and integral equations are obtained by proving some theorems.
- 5) The obtained results can be included into the reference mathematical library and implemented into Mathematics program products, MathCAD, Math Lab means. It can be useful for scientists, engineers, mathematics at solving the different problems.

## Bibliography

- [1] Aronzajn A. Theory of reproducing kernels // Trans. American Math. Soc., 68:337-404, 1950.
- [2] *Functions with double orthogonality in radio electronics and optician* / By edition M.K. Razmahnin and B.P. Yakovlev. M.: Sov. radio, 1971. 256p. (in Russian).
- [3] Saitoh S. Integral Transforms, Reproducing Kernels and Their Applications, Pitman Res. Notes in Math. Series, 369 (1997), Addison Wesley Longman, UK.
- [4] Laslo Mate. Hilbert Space Methods in Science and Engineering, 1990. 200 p.



- 
- [5] *Daubeshies I.* Ten lectures on wavelets. CBMS-NSF. Regional conference series in applied mathematics. SIAM, Philadelphia, 1992.
- [6] *Sethu Vijayakumar, Hidemitsu Ogawa.* RKHS based Functional Analysis for Exact Incremental Learning / Neurocomputing: Special Issue on Theoretical analysis of real valued function classes, Vol.29, No.1-3, pp.85-113, Elsevier Science (1999). [http://www-clmc.edu/sethu/research\\_detail.1html](http://www-clmc.edu/sethu/research_detail.1html).
- [7] *Saitoh S.* New Norm Type Inequalities for linear mappings // *Jornal of Inequalities in Pure and Applied Mathematics.* Victoria Univ., 2003. Volume 4. Issue 3. Article 57. <http://jipam.vu.edu.au>.
- [8] *Girosy F.* An equivalence between sparse approximation and support vector machines // *Neural Computation,* 10(6):155-1480, 1998.
- [9] *Vapnik V.* The nature of Statistical Learning Theory. Springer-Verlag. New York, 1995.
- [10] *Chumachenko S.V.* Summation method of selected series for IP-core design // *Radioelectronics and Informatics.* 2003. N3. C. 197-203 (in Russian).
- [11] *Chumachenko S.V.* Theorem about some integral identities based on Series Summation Method in RKHS // *Radioelectronics and Informatics.* 2004. №1. C. 113-115 (in Russian).
- [12] *Chumachenko S.V.* Reproducing kernel Hilbert spaces and some theirs applications // *Radioelectronics and Informatics.* 2003. № 4. C. 141-144 (in Russian).
- [13] *Tranter C.J.* Integral Transforms in Mathematical Physics. London: Methuen & Co. Ltd., New York: John Wiley & Sons, Inc. 1951.
- [14] *Doetsch G.* Anleitung Zum Praktischen Gebrauch Der Laplace-transformation und Der Z-transformation. R. Oldenbourg. Munhen, Wien, 1967.
- [15] *Angot A.* Complements de Mathematiques. A L'usage Des ingenieurs de L'elektrotechnique et Des Telecommunications. Paris, 1957.
- [16] *Watson G.N.* A Treatise on the Theory of Bessel Functions. Cambridge: Cambridge University Press. 1952.
- [17] *Titchmarch E.C.* The theory of functions. Oxford Univ. Press, 1939.
- [18] *Ahiezer N.I.* The lectures of the approximating theory. Moscow: Nauka, 1965. 408p. (in Russian).
- [19] *Chumachenko S.V., Govhar Malik, Imran Saif Chattha.* Series Summation Method in HSRK // *Proc. of the international Conference TCSET'2004 "Modern Problems of Radio Engineering Telecommunications and Computer Science".* February 24-28. 2004. Lviv-Slavsko, Ukraine. P.248-250.
- [20] *Chumachenko S.V.* Solving Electrodynamics Problems by Reproducing Transformations Method // *Proc. BEC 2004.* Tallinn. October 3-6, 2004. PP. 319-322.
- [21] *Chumachenko S.V., Gowher Malik, Khawar Parvez.* Reproducing Kernel Hilbert Space Methods FOR CAD Tools // *Proc. EWDTW, 2004.* Yalta-Alushta, September 23-26. PP. 247-250.
- [22] *Svetlana Chumachenko, Vladimir Hahanov.* Reproducing Transformations method for IP-core of summatory and integral equations solving // *Proc. of DSD 2004 Euromicro Symposium on Digital System Design: Architectures, Methods and Tools.* August 31 - September 3, 2004, Rennes – France (Work in progress).
- [23] *Titchmarch E.C.* The theory of functions. Oxford Univ. Press, 1939.
- [24] *Prudnikov A.P., Bryichkov Yu.A., Marichev O.I.* Integrals and series. Moscow: Nauka, 1981. 800p. (in Russian).

---

### Authors' Information

**Chumachenko Svetlana** – Ph. D. Senior Scientist and a professor assistant. Address: Kharkov National University of Radio Electronics, Ukraine, 61166, Kharkov, Lenin Avenue, 14, Phone: (+380)-57-7021-326, e-mail: [ri@kture.kharkov.ua](mailto:ri@kture.kharkov.ua).

**Kirichenko Ludmila** – Ph. D. Senior Scientist and a professor assistant. Address: Kharkov National University of Radio Electronics, Ukraine, 61166, Kharkov, Lenin Avenue, 14, Phone: (+380)-57-7021-335, e-mail: [ludmila@kture.kharkov.ua](mailto:ludmila@kture.kharkov.ua).

## EXAMINATION OF ARCHIVED OBJECTS' SIZE INFLUENCE ON THE INFORMATION SECURITY WHEN COMPRESSION METHODS ARE APPLIED

Dimitrina Polimirova–Nickolova, Eugene Nickolov

**Abstract:** After giving definitions for some basic notions as risk and information security, archived objects with different size as well as different compression methods are examined and described. An experiment is made using different compression methods with different objects' size and type, followed by an analysis and an evaluation of the obtained results. In the end, some conclusions and recommendations for future work are suggested.

**Keywords:** Archived Objects, Archiving Programs, Information Security, Compressed Objects, Methods of Compression, Password, File Extensions.

---

### Introduction

The new technologies' development extends the necessity of creation and use of archived objects, especially when they are protected by differing in length passwords depending on the needs and the possibilities. The trend that dominates in their use is the necessity of creation of real time working high-speed and effective compression of information flows immediately after their coming. That's why the examination and the analysis of different compression methods and their varieties become an exceptionally pressing problem which has found different solutions in the past decades.

Archived objects with different length (from 1Mb to 128Mb) are examined and analyzed here, on which different compression methods are applied using an 8-digit fixed password, comprising for simplicity the numbers from 0 to 7 in ascending and consecutive order. We shall consider as archived objects these objects which contain data, usually saved in file, reserved for a later use. We shall consider as compressed objects these objects, which length is decreased in order to limit the costs for their saving and transportation and to increase their information security. We shall understand information security as risk evaluation which is a correlation among threats, vulnerabilities, losses and undertaken contra measures. This is an integral evaluation giving a descriptive non-mathematical interpretation of the most important factors that participate in these processes. We shall understand the risk as probability of realizing a specific attack on a specific object with a specific importance (cost). We shall understand as threats the possible attack scenarios realizing particular actions. We shall understand as vulnerabilities the possible breaches (conscious, unconscious, accidental, no-accidental) in the protection of a specific object. We shall understand as losses either a numeric expression of damages caused by the attack (in monetary equivalent for appropriate currency), or a specific state of an object (which could not be evaluated in a numerical way). We shall understand as contra measures the set of all preliminary measures for counteraction, all measures for reaction during the attack and all measures for removing the consequences.

In this study the archived objects are represented by 6 extension types, appertaining to 6 major types of archiving programs. They are chosen from more than 300 archiving programs, known up to now. They are:

- 1) E-mail archiving programs – this type of archiving programs use the relative uniformity of information flow (e-mail traffic), to select the most suitable compression methods.
- 2) Converting archiving programs – these archiving programs are able to convert objects compressed by a specific method in objects compressed by another method.
- 3) Multiple archiving programs – these programs perform some consecutive archiving operations on the different parts of an object using a few compression methods with different properties.
- 4) Image archiving programs – they are of prime importance in the modern real time processing of video and image web-objects. The trend dominating in this processing is the compulsory immediate compression of the object after its creation. The object transmission and processing are performed completely in compressed form up to the final moment of its reproduction by the relevant media.

5) Data archiving programs – these programs are specialized in the processing and use of compressed objects, obtained from information flows which could be characterized as “data” (in this case we are interested by the fact that in the different phases of their existence data pass in compressed form, exist some time in this “stored” state and are decompressed after that).

6) Executable archiving programs – these programs look for and obtain some compression specificity, connected with the possibilities for running the compressed objects.

## The Experiment

The objects of examination are compression methods (different types of compressing programs) focusing on the initial object length and its influence on the compression process [1, 2].

The following tasks are posed in connection with this study:

1) To select suitable objects, that are enough to draw conclusions valid for most file extensions and relevant compression methods (Table 1).

**Table 1**

EXTENSION	PROGRAM / INFORMATION	TYPES OF ARCHIVING PROGRAM
.DBX	Outlook Express Email Folder	E-mail archiving programs
.GZIP	GNU Zip Compressed Archive	Converting archiving programs
.TAR	Tape Archive File	Multiple archiving programs
.CPT	Corel Photo-Paint Image (Corel)	Image archiving programs
.DOC	Word Document (Microsoft)	Data archiving programs
.EXE	Executable file (Microsoft)	Executable archiving programs

2) To select the basic compression methods and their varieties which will be applied on the selected objects. Based on previously carried out experiments, 3 basic types of compression methods are selected: ACE, RAR, ZIP. The following versions of the selected compression programs are used in the experiments:

WinACE 2.6b5;

WinRAR 3.42;

WinZIP 9.0 (6028).

In this study they will be applied with their default settings (Compression level: Normal, with 8-digit fixed password) [3].

3) To identify the initial size of objects to be compressed for the different file extensions. The range of the selected initial object sizes – 1Mb, 4Mb, 8Mb, 16Mb and 32Mb is large enough to draw conclusions. Partial examinations are made for objects with 64Mb and 128Mb sizes.

4) To compress objects selected according to conditions set in tasks one and three by the compression methods, described in task two.

**Figure 1a Compressed 1Mb files**

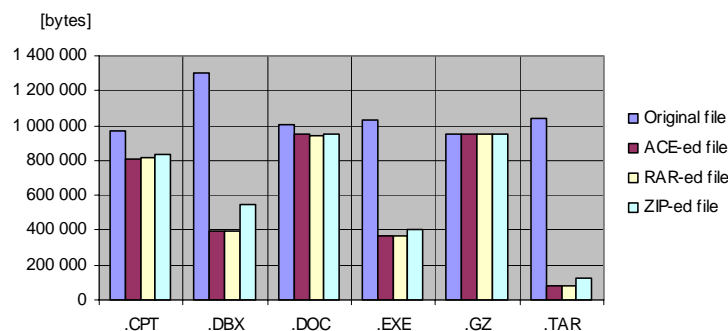


Figure 1b Compressed 4Mb files

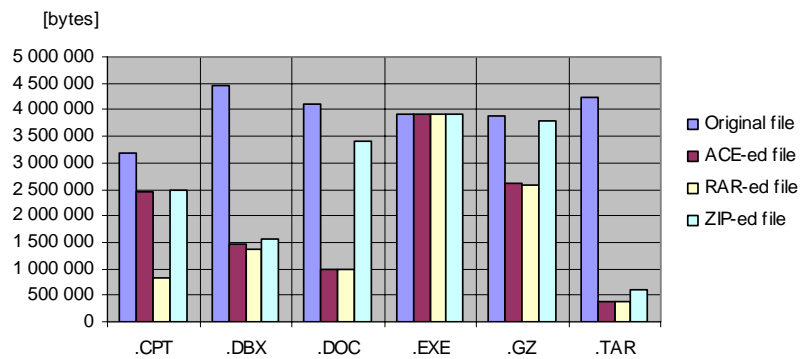


Figure 1c Compressed 8Mb files

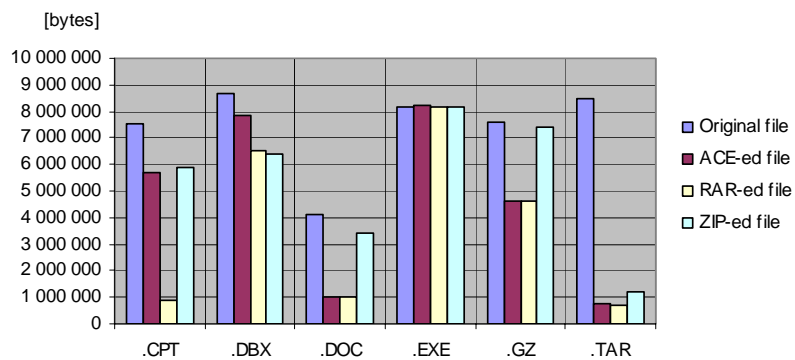


Figure 1d Compressed 16Mb files

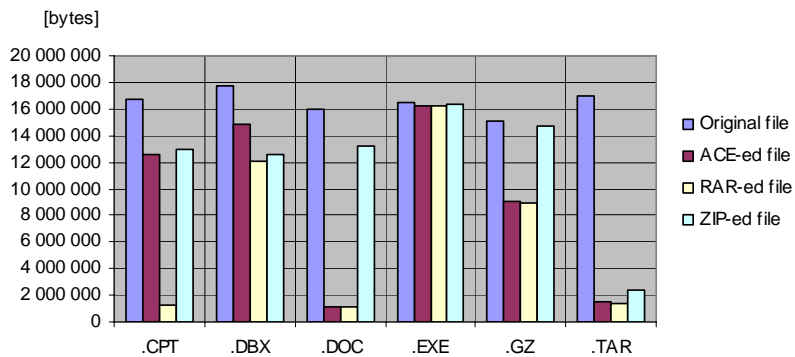
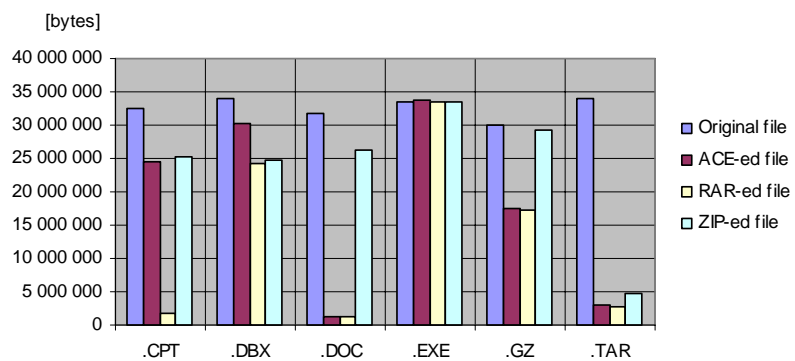


Figure 1e Compressed 32Mb files



The obtained results are illustrated on Figure 1a-1e.

5) To juxtapose the sizes of the initial and obtained after compression file objects. Their percentage decrease or increase is illustrated on Figure 2a-2f.

Figure 2a

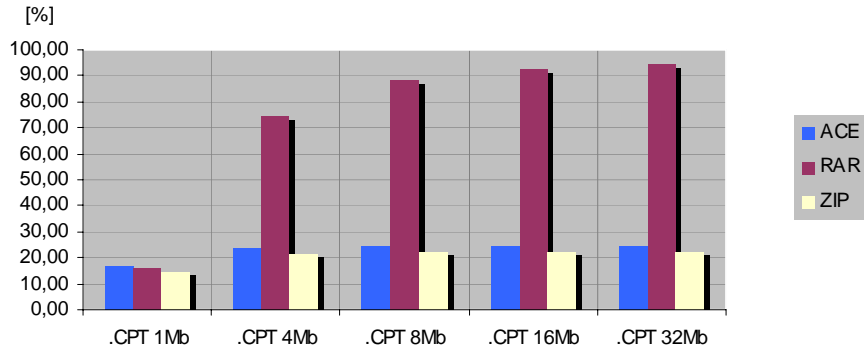


Figure 2b

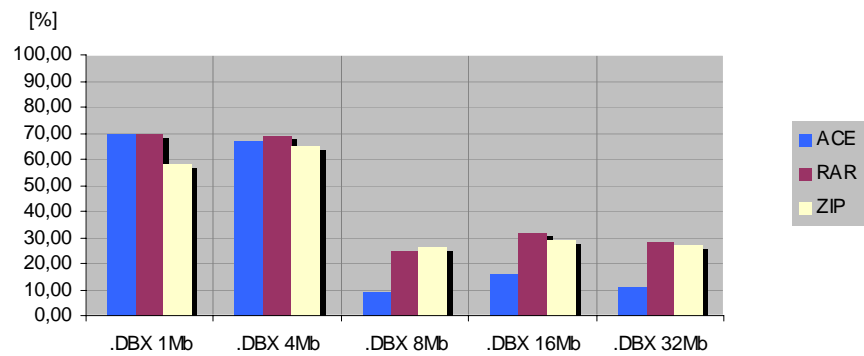


Figure 2c

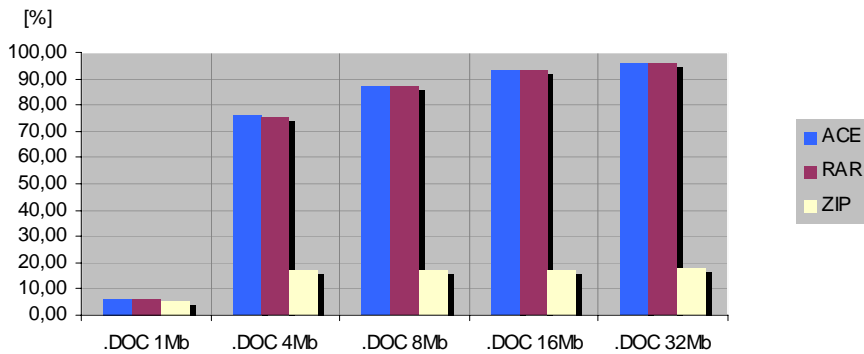


Figure 2d

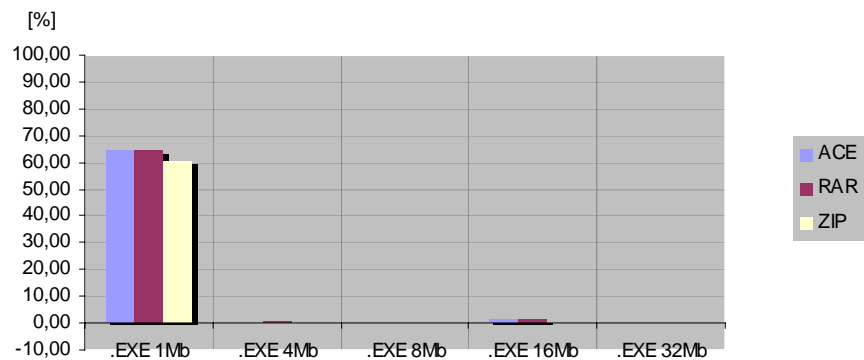


Figure 2e

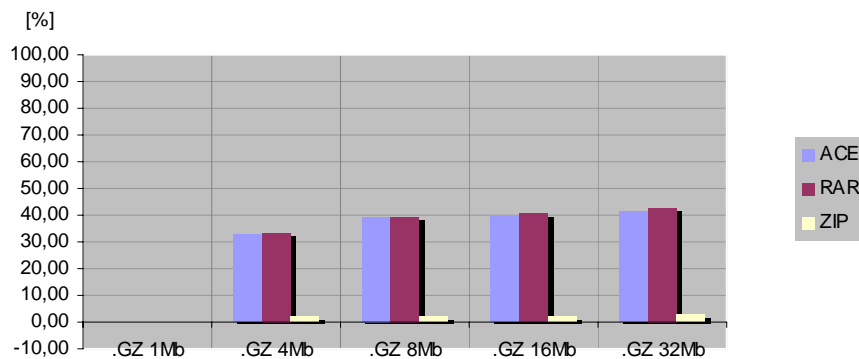
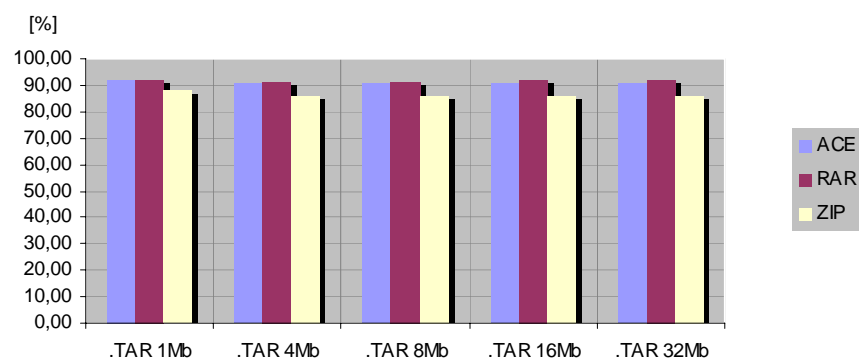


Figure 2f



The following assessments could be made from the experiments which were carried out:

- 1) The selected objects perform successfully the tasks that were posed. The assumptions that were made do not influence the obtained results. The selected 6 from 300 extensions are enough to make the necessary generalization.
- 2) The selected compression methods guarantee the fulfillment of the tasks posed. The experiments that were carried out could be generalized for other methods.
- 3) The initial object size is well-chosen. The experiments with sizes 1Mb, 4 Mb, 8 Mb, 16 Mb and 32 Mb show no need to carry out experiments with sizes 64 Mb and 128 Mb, because there are no qualitative changes in the trends, and the experiment costs grow up sharply for these values.
- 4) The use of the Normal compression level allows in most cases generalization of the obtained results for the other higher or lower compression levels. The obtained results show the existence of local extremes for specific values which sets the pattern for future examinations.
- 5) Best results with respect to the size are obtained for .DOC file with 32Mb length, compressed with the ACE compression method. Poorest results with respect to the size are obtained for .EXE file with 8Mb length, compressed with the ACE compression method.
- 6) The use of ACE and RAR compression methods give in most cases comparable results, by contrast with the ZIP compression method, which gives significantly differing results. These assessments are valid for the specific program versions listed above.
- 7) The used fixed password, described above is sufficient for an initial level of information security in the frame of the experiments which were carried out. The obtained results allow to make the assumption that additional examinations should be made with respect to the password length and its structure as well. It is necessary to select a standardized method of information security evaluation for similar objects and password differing on size and structure.

8) The largest difference in the sizes of the initial and the resulting objects is observed between 1Mb and 4Mb. The obtained results show flowing changes for the next levels. This is valid for the three selected compression methods.

---

### Conclusion

The examinations that are made and the experiments that are carried out show the large prospects of such file object processing because beside the size change it results in information security increase.

The initial object size exercises a significant influence when specific compression methods are applied.

The largest initial size does not always set the pattern for a largest compression level and a largest information security, accordingly.

The implementation of the appropriate level of information security and risk evaluation in information flow processing is connected to the analysis of the archived objects' size.

---

### Bibliography

[1] David Salomon, Data Compression: The Complete Reference, Springer Verlag New York, Inc., 2004.

[2] Alistair Moffat, Andrew Turpin, Compression and Coding Algorithms, Kluwer Academic Publishers, 2002

[3] Rob Shimonski, Introduction to Password Cracking, IBM Developer Works Hacking Techniques article, 2002.

---

### Authors' Information

**Dimitrina Polimirova-Nickolova** – PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, E-mail: polimira@nlcv.bas.bg.

**Eugene Nickolov** – Prof., DSc, PhD, Eng, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, E-mail: eugene@nlcv.bas.bg.

## ВИЗУАЛИЗАЦИЯ НА АЛГОРИТМИ И СТРУКТУРИ ОТ ДАННИ

**Ивайло Петков, Сергей Георгиев**

**Анотация:** В статията се разглежда, web-базирано моделиране на абстрактни структури от данни и прилагането на избрани алгоритми върху тях. С цел използването на системата от по-широк кръг потребители е реализиран адаптивен потребителски интерфейс.

**Keywords:** граф, двоично дърво на наредба, Дийкстра, Форд-Белман, цикъл

---

### Въведение

Графите и дърветата са неразделна част от курсовете по структури от данни. Тяхната значимост се определя от широкото им практическо приложение. Голямото количество дадена информация изисква умело структуриране с цел бърз достъп до определена част от нея. Съществуват различни структури от данни, които се използват за това, но поради своята гъвкавост и бърз достъп до всеки елемент дърветата са една от най-ефективните и широко разпространени. Графите са мощен инструмент при моделирането на много задачи от различни области на науката и практиката. Изборът и прилагането на подходящ алгоритъм върху граф дава ефективно решение на редица практически проблеми.

Самите структури дърво и граф, както и алгоритмите върху тях, изискват добро пространствено въображение и склонност към абстрактно мислене. Всичко това поставя преподавателите пред трудна и отговорна задача, която би могла успешно да бъде подпомогната с помощта на визуализация.

В настоящата статия е разгледана web базирана система GM&TD(<http://www.edusoft.fmi.shu-bg.net>), даваща възможност за лесно и интуитивно конструиране на графи и дървета и демонстрация на някои широко приложими алгоритми.

### Предпоставка

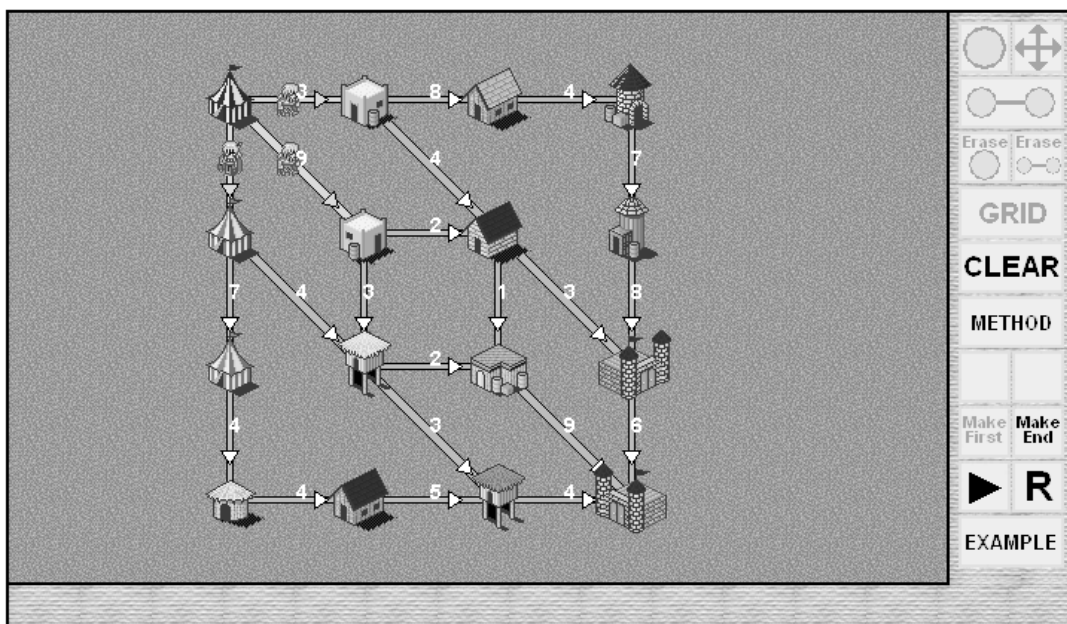
В момента има много сайтове, предлагащи демонстрация на алгоритми върху абстрактни структури от данни. След анализ на предлаганите от тях услуги могат да бъдат отбелязани следните недостатъци:

- Липсва реализация на български език [8],[9],[10].
- Системите, работещи с графи демонстрират само един алгоритъм и липсва възможност за сравнителен анализ на различни алгоритми върху един и същ граф. Например системата на Carla Laffra [8] демонстрира само един от най-известните алгоритми върху графи - алгоритъма на Дийкстра. За проиграването на друг алгоритъм върху същия граф трябва да се търси друга система.
- Реализацията на основните операции върху дървета не е съпроводена с достатъчно ясни обяснения [9].
- Системите, визуализиращи дървета, не разполагат с възможност за избор на поддърво за търсене на заместник при изтриване на елемент с два наследника [9],[10].
- Повечето системи, обработващи дървета, не предоставят възможност за постъпково изпълнение на алгоритмите и контрол на скоростта на анимацията.

### Същност

Системата GM&TD(<http://www.edusoft.fmi.shu-bg.net>) се състои от два модула – Graph Modeling и Tree Demonstration.

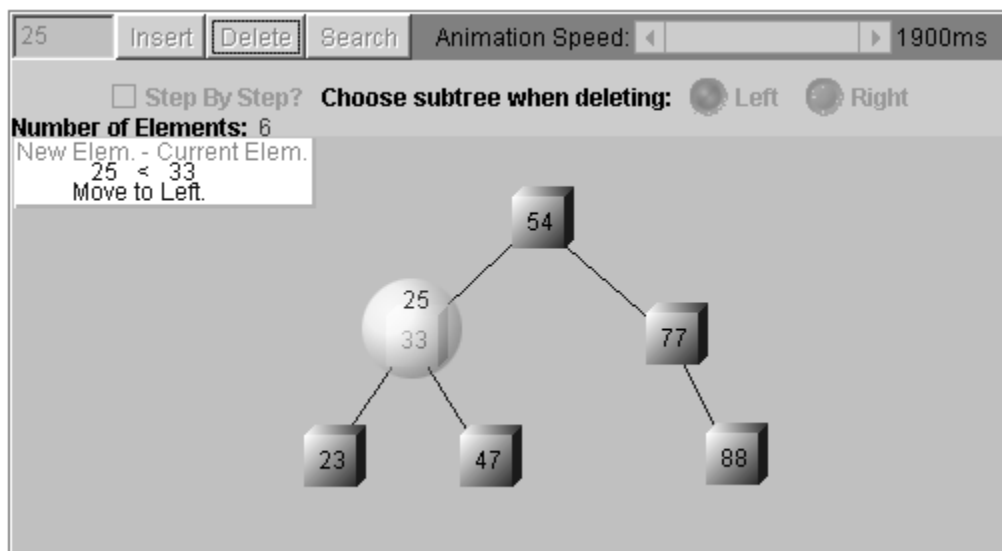
**Graph Modeling** (фиг.1) е удобен инструмент за визуализация на графи. Интерфейсът на модула е реализиран на български и английски език. Визуализира алгоритмите за обхождане на граф в ширина и дълбочина, алгоритмите на Дийкстра, Форд-Белман за търсене на най-кратък път между два върха и алгоритъм за намиране на цикли в графа. Модулът притежава интерфейс, който позволява на потребителя да създава свои собствени графи или да избира готови образци, върху които да експериментира начина на действие на изброените алгоритми.



Фиг.1



**Tree Demonstration** (фиг.2) визуализира алгоритмите за вмъкване, изтриване и търсене на елемент в двоични дървета на наредба. Интерфейсът на модула е реализиран на български и английски език. Предоставя вградено специализирано поле, съдържащо обосновка за всяка стъпка от съответния алгоритъм. Други подобрения са лентата за скролиране, чрез която потребителят има възможност да избира времето за изчакване между две последователни стъпки от избрания алгоритъм, както и checkbox, с чиято помощ може да преминава към следваща стъпка само чрез натискане на специален бутон.



Фиг.2

Изграждането на GM&TD като web система се налага по две причини:

1. Системата е предназначена за обучение и като такава трябва да бъде независима от платформата.
2. Системата трябва да бъде достъпна за всички потребители, интересувани се от поставения проблем независимо от тяхната възраст, местоположение или цел.

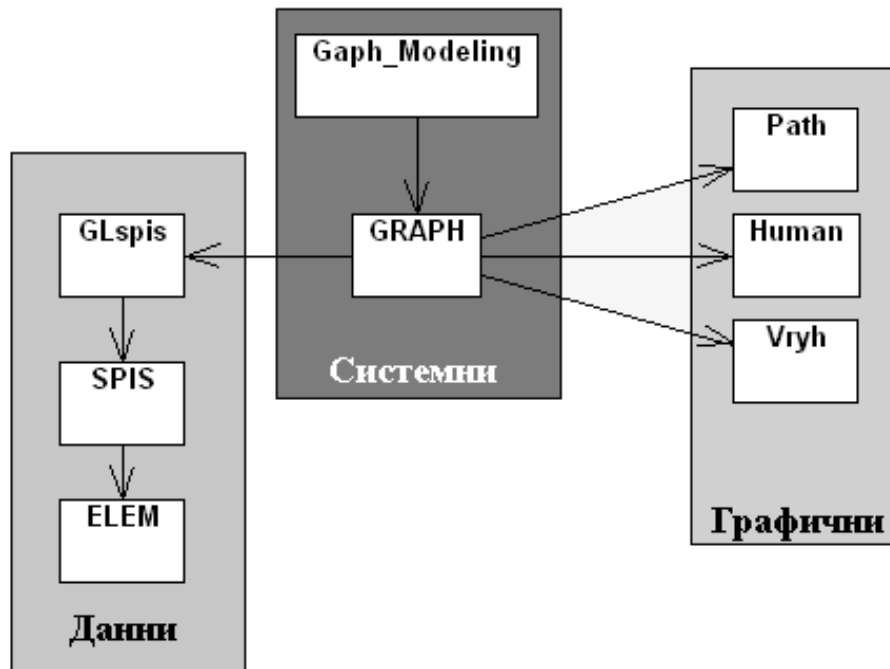
За да отговорят на всички наложени им изисквания системата е разработена на програмния език Java [4]. Java е прост, обектно-ориентиран, защитен от грешки, сигурен, преносим, независим от платформата, интерпретируем, паралелен и динамичен език. Програмният код на Java се компилира до преносим междинен код (байткод), който изисква интерпретатор, за да бъде изпълнен.

**Graph Modeling** е изградена от осем класа (фиг. 3), които се разделят условно на три групи: системни, графични и класове, предназначени за складиране на данни.

- Системните класове Applet и GRAPH разполагат с основните методи и играят главно свързваща роля. Класът Applet е основният за системата и се извиква от браузъра като му се предава управлението. След инициализирането на декларираните обекти в него се създава нишка която опреснява екрана на всеки сто хилядни от секундата. Въпреки че опресняването се извършва сравнително малко пъти в секундата за да се избегнат премигванията на екрана се използва широко разпространената в графичното програмиране техника **Double Buffering**. В GRAPH класа се намират реализациите на методи за: добавяне и премахване на връх, преместване на връх на произволна позиция в работната област, добавяне и премахване на ребро, изчистване на работната област, грид (мрежа от точки помагача по-лесното разполагане на върховете), избиране на начален и краен връх, реализациите на всички методи който демонстрира системата.
- Графичните класове Path, Vgryh съдържат набор от методи, реализиращи единствено различните начини за визуализация на ребрата и върховете в графа. Класа Human реализира всички възможности на човека, което обхожда графа. Този клас е реализиран с оптимални алгоритми,

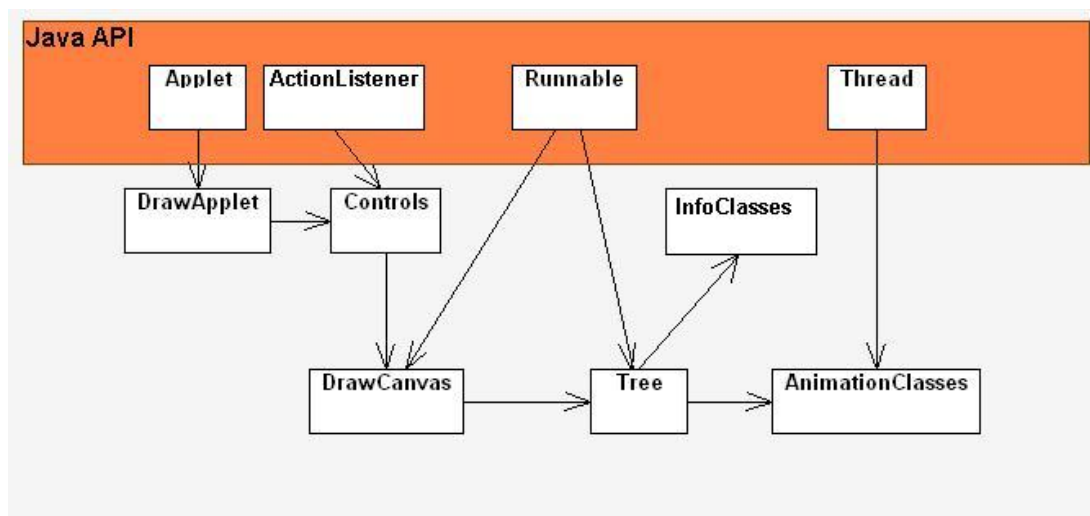
които по никакъв начин да не забавят визуализацията. За разлика от тях Menu класът е много по-функционален. Той обработва съобщенията от мишката, променя външния вид на менюто в зависимост от действията на потребителя, определя какъв метод трябва да бъде извикан. Всъщност визуализацията се осъществява по нива. Applet класът извиква метода за изчертаване на GRAPH, от своя страна той извиква методите за изчертаване на класовете Path, Vryh и Menu.

- Класовете, предназначени за складиране на данни, реализират обща структура от данни наподобяваща списък от списъци. В повечето подобни системи се използват предварително заделени квадратни матрици на съседство. При по-малки размери на графа матрицата съществува в паметта в пълния си определен предварително размер, което излишно утежнява системата.



Фиг. 3

**Tree Demonstration**, както беше споменато по-горе, се състои от 3 подсистеми, които от своя страна са изградени от класове (фиг.4), разделени в зависимост от своята функционалност в следните подгрупи: реализиращи интерфейса, анимирането на алгоритмите, помощни и базови класове.



Фиг.4

- Базовите класове DrawApplet и Tree, подобно на системните в **Graph Modeling**, разполагат с основните методи и играят главно свързваща роля. DrawApplet наследява класа Applet от Java API, което е продиктувано от това, че системата е реализирана като аplet. В този клас се създават работната област и панелът с контролите. Съответно тук се задават и основните им стойности като размери и координати на заеманите от тях области. Класът Tree създава структурата дърво. Съдържа методи, които осъществяват алгоритмите за вмъкване, търсене, изтриване на елемент и отпечатване на данните.
- DrawCanvas и Controls са класовете, организиращи интерфейса на приложението. Първият представлява работната област, в която се визуализира генерираното от потребителя дърво. Заради наличието на анимирани ефекти тук се наложи да се въведат и техники като горепосочената **Double Buffering**, **MediaTracker** и други. Controls е отговорен за разполагането и оразмеряването на контролите. И двата споменати тук класове притежават методи, които обработват съобщения, настъпващи от мишката (mouseDown, mouseUp и mouseDrag).
- AnimationClasses са набор от класове, отговорни за анимирането на визуализацията. За тази цел в тях са използвани техники взаимствани от растерната графика като растеризация на отсечка и на окръжност. AnimationClasses са наследници на Thread от Java API и съответно за всеки един ефект, който трябва да се изобрази в работната област се пуска отделна нишка към процесора.
- Помощните класове InfoClasses капсулират информация за елементите на дървото. Тази информация е както следва: ключ на елемент, координати за изобразяване в работната област, фонен цвят и други.

Системният интерфейс на **Graph Modeling** е изграден от тринадесет бутона, разделени в две категории според своето предназначение. Базовите бутони са свързани с основните методи, необходими за изграждането и разполагането на графа в работната област на системата. Функционалните бутони служат за избор на алгоритъм и неговото управление по време на визуализация.

- **Базови бутони**



Поставяне на нов връх. Достатъчно е да натиснете бутона на мишката на произволно място в работната област на системата. Върхът се именува автоматично.



Промяна позицията на връх. Необходимо е единствено да задържите бутона на мишката върху върха, след което да го преместите на желаната позиция с последващо отпускане на мишката.



Създаване на ребро между два върха. Последователно се маркират двата върха които ще свързвате, появява се стрелка и поле, в което трябва да въведете теглото на реброто.



Изтриване на връх. Натиснете бутона на мишката върху върха, който искате да бъде изтрит. Върхът се изтрива заедно с всички свързани с него ребра.



Изтриване на ребро. Последователно се маркират двата върха, между които ще се изтрива реброто. Последователността се определя от посоката на реброто.

**GRID**

Активиране на помощната мрежа.

**CLEAR**

Изчистване на работната област.



Определя върха, от който трябва да започне изпълнението на алгоритъма.



Определя върха, в който трябва да приключи изпълнението на алгоритъма.

**EXAMPLE**

Въвежда в работната среда готови графи за изпробване на алгоритмите.

**Функционални бутони****METHOD**

Отваря помощно меню, от което избирате алгоритъма, който трябва да бъде визуализиран.

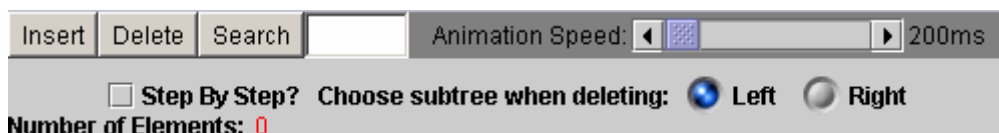


Извиква изпълнението на следващата стъпка от алгоритъма.



Рестартира алгоритъма. Алгоритъмът започва своята работа от първата стъпка.

В системата **Tree Demonstration** се залага на изчистения от страна на контроли интерфейс, за да се улесни работата с приложението. По време на изпълнение на даден алгоритъм, в работната област се изобразява както генерираното дърво, така и допълнителна информация, описваща всяка стъпка от протичането му. Чрез промяна цвета на съответния елемент се насочва вниманието на потребителя към мястото в дървото, до което е стигнал алгоритъмът. Точно за това се наложи вмъкването на контрола, с която е възможна настройка на скоростта на анимацията според личните желания и възможности на съответния човек.



Фиг.5

Включени са следните контроли (фиг.5):

	вмъкване на елемент
	изтриване на елемент
	търсене на елемент
	поле за въвеждане на елемент, върху който ще се изпълнява желаната операция. След въвеждане на елемента в това поле се избира операцията чрез бутоните Insert, Delete, Search.
	радио-бутони за избор на елемент, заменящ елемента за изтриване. Left – заменя елемента за изтриване с най-десния от лявото поддърво, а Right – с най-левия от дясното поддърво.
	лента за скролиране, даваща възможност за контролиране скоростта на анимацията
<input type="checkbox"/> Step By Step?	checkbox за избор на постъпково изпълнение на алгоритъм

**Заклучение**

В настоящата разработка беше разгледана система за визуализация на структури от данни. Тя тласка потребителя към едно по-творческо мислене. Мотивира го да експериментира със системата и да придобие професионален интерес към предметната област. Разработената система би могла да бъде използвана като помощно средство в обучението на студенти и ученици в курсовете, изучаващи дървета и графи. Работи се и върху нейното обновяване с нови алгоритми и структури от данни.

---

## Литература

---

1. Амералд Л. Алгоритми и структури от данни в C++. – ИК СОФТЕХ София, 2001.
2. Седжуик Р. Алгоритми в С. – Софтпрес, 2002.
3. Уирт Н. Алгоритми + структури от данни = Програми.
4. Стоун К. и Д. Уебър. Java 2: Програмиране за Интернет. – LIO Book Publishing, 2000.
5. Георгиев И., Г. Герасимов, Ц. Врабчарска, И. Сиклунов и Л. Матеев. Графично програмиране.
6. Ламот А., Д. Ратклиф, М. Семинейтър, Д. Тайлър. Програмиране на съвременни компютърни игри. – НИСОФТ, София 1996.
7. Мирчев И. Графи - Унив. изд. Неофит Рилски Благоевград, 2001
8. <http://www.dgp.toronto.edu/people/JamesStewart/270/9798s/Laffra/DijkstraApplet.html>
9. <http://ww3.java3.datastructures.net/animations.html>
10. <http://www.student.seas.gwu.edu/~idsv/idsv.html>

---

## Информация за авторите

---

**Ивайло Петков** – студент в Ш.У. “Епископ Константин Преславски”, е-mail: [i\\_petkow@yahoo.com](mailto:i_petkow@yahoo.com)

**Сергей Георгиев** - студент в Ш.У. “Епископ Константин Преславски”, е-mail: [s\\_ggeorgiev@yahoo.com](mailto:s_ggeorgiev@yahoo.com)

## НОВ ПОДХОД КЪМ КОНСТРУИРАНЕТО НА БЛОК-СХЕМНИ ЕЗИЦИ

Стоян Порязов

**Резюме:** На определени етапи от анализирането и проектирането на сложни системи, блок-схемните езици имат редица предимства в сравнение с текстовите. Въпреки това, в информатиката те имат по-малко приложение, отколкото в електрониката, хидравликата и други области, където са задължителни. За това има няколко причини, по-важните от които са разгледани в статията. В статията е предложен нов подход към конструирането на блок-схемни езици в областта на информационното моделиране.

**Ключови думи:** блок-схемни езици, графични знаци, информационно моделиране.

**Основната парадигма** е, че с предлаганият език се описва модел на структурата, състоянието и функционирането на информационна машина, намираща се в неизвестна за нея външна среда. Описваната машина може да бъде разглеждана и като информационен модел на някаква реална или въображаема система.

**Същини:** Същините (същност, англ. "entity") се състоят от едно или повече свойства, между които има съответствия, релации и връзки.

**Семантичните елементи** в графичния език са свойствата и състоянията на моделираните същини и операциите с тях. Едно от базовите свойства е "носител на знак".

**Базовите видове операции, свойства и същини, за които има предвидени графични знаци, са:**

1. **Указване.** Използва различни видове указатели (съответствия).

2. **Преструктуриране.** Включва операциите създаване (копиране, генериране) и унищожаване на единични свойства, както и на съчетания от свойства, осъществени посредством връзките: канал, верига (двупосочен канал), припокриване и композиция.
3. **Наблюдение.** Въз основа на наблюдаваните същности (вътрешни, или външни за информационната машина), създава знаци.
4. **Интерпретация.** Въз основа на знаци, генерира вътрешни, или външни същности и свойства. Например, управлява потоците от същности (комутация (превключване) на канали).
5. **Пренасяне.** Реализира се посредством канали, които могат да се комутират, и припокриване.
6. **Съхранение.** Операциите с памет могат да бъдат: вход, изход, четене (копиране), търсене на обект, проверка за налично място, проверка за обект на посочено място.
7. **Пресмятане.** Работа (създаване, промяна, унищожаване) със стойности на съществуващи същности. Може да се използват различни теории, използващи стойности - булеви, логически, числени и т.н.
8. **Забавяне.** Всички операции се считат за извършвани за нулево време. За да се моделира забавянето в реалните системи, е предвиден специален знак "таймер".
9. **Въздействие/реакция.** Необходими са, защото има случаи, при които резултатът (реакцията) настъпва известно време след прилагане на въздействието.
10. **Защита** (охрана). Например, защита от вируси, прегряване и т.н.
11. **Поддръжка** (възстановяване). Например, опресняване на компютърната памет.
12. **Управление.** Може да се разглежда като мета-информационна машина, която е необходима за работата на основната (целевата). Използват се знаците за вече изброените операции, като само за управленските потоци са предвидени специални знаци.

Очевидно изброените операции не са елементарни, нещо повече, някои не са и информационни, но всички те са необходими за описание на реалните телекомуникационни и компютърни системи (както и за повечето други) и затова са включени в списъка.

**Графични блокове:** Основата за създаване на блоковете се състои от четири точки, намиращи се във върховете на правоъгълник. Размерите на правоъгълника се определят от конкретните нужди, за всеки отделен блок. Страните на правоъгълника се затварят от графични знаци. Основните значения на блоковете са: същност, множество, пояснение, (виртуално) устройство и операция. Те се определят от създаващите блока графични знаци и от взаимното положение на блоковете.

**Графични знаци:** Графичните знаци се състоят от една или повече линии. За да се запази ориентацията на знака относно вътрешната част на блока се използват геометричните трансформации "ротация" и "осева симетрия" на графичните представяния. Всеки графичен знак има основно значение, което се запазва при всичките му употреби, но се допълва и уточнява в зависимост от положението на знака в съответния блок и от значението на блока.

Описаният подход позволява създаването на голям брой блокове въз основа на малък брой графични знаци, което го прави лесен за запомняне и мощен като моделиращ инструмент. Той е изпробван успешно при моделиране на телекомуникационни системи и е в процес изпробването му при моделиране на бизнес-системи.

---

### Информация за автора

---

Стоян Порязов – Институт по математика и информатика, БАН, ул. "акад. Г. Бончев", блок 8

---

## СИМУЛАЦИЯ НА НЯКОИ ОСНОВНИ ГРЕШКИ ПРИ ИЗМЕРВАНЕ НА ОПТИЧНИ ПАРАМЕТРИ ЧРЕЗ ПРИЛАГАНЕ НА ОБРАТНА ЗАДАЧА В ОПТИКАТА И ФАЗОВО-СТЪПКОВ МЕТОД

Георги Стоилов

**Абстракт:** За решаване на обратната задача в оптиката е използвано обратно Фурие преобразуване на изображение, получено от преминаването на лазерен сноп през обекта. Предложено е използването на фазово-стъпков метод за намиране на разпределението на фазата. Това позволява изчисляването на коефициента на пропускане в комплексен вид за всяка точка на изследвания обект. Използването на опорно измерване елиминира влиянието на параметрите на измервателната система.

Показан е основният алгоритъм за изчисляване. Направена е компютърна симулация на влиянието на по-важните параметри на оптичната схема върху точността на измерване. Симулирани са грешки при позициониране на основните оптични елементи и неточно фазово отместване. Анализирани е използването на АЦП с различна разрешаваща способност. Дискутирани са условията и ограниченията за успешно прилагане на метода.

Този метод може да бъде използван за измерване и окачествяване на малки и микробекти.

**Ключови думи:** 2D-оптични параметри, обратна задача в оптиката, фазово-стъпков метод, компютърна симулация

---

### Въведение

В хода на провеждане на експеримент за измерване на оптичните параметри на малки и микробекти чрез решаване на обратната задача в оптиката [1] т.е. като се знае функцията на входящата и преминалата вълна да се намери предавателната функция на един обект, се наложи класифицирането на факторите, които внасят грешка при измерването и оценка на допустимите им стойности и точност. Интерферометричното естество на метода и използването на цифрови камери показва, че трябва да се направи оценка и съгласуване на параметрите на интерференчната картина и тези на камерата. При оптични измервания във Фраунhoferовата зона образът, който се получава при обекти с периодична структура, е концентриран в няколко области в зависимост от пространствените честоти, присъщи на обекта. Измерването на неперидични обекти налага измерването на значително по-слаби сигнали и прави важна оценката за точността, с която се снимат кадрите от камерите.

---

### Метод на измерване

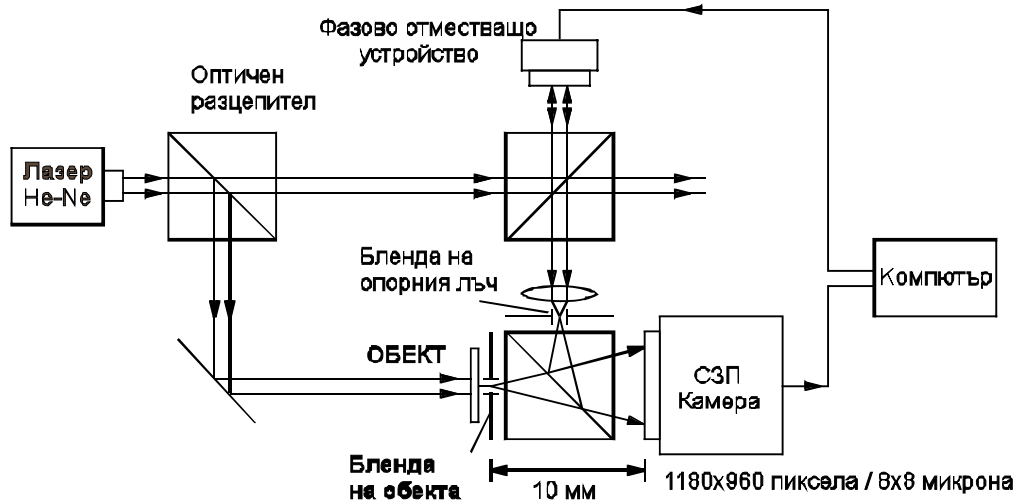
Методът на измерване се базира на решаване на обратната задача в оптиката – предавателната функция на обекта се изчислява от функцията на осветяващата вълна и функцията на вълната, измерена след обекта. Фазово-стъпковият метод [2] дава възможност за измерване на амплитудата на вълната в комплексен вид. Едновременното му използване и измерването на интензитета на дифракциралата светлина прави възможно изчисляването на коефициента на пропускане в комплексен вид. За да се отстрани влиянието на непознатите параметри на осветяващата система, се прави допълнително опорно измерване.

---

### Експериментална постановка

Схемата на експерименталната постановка е показана на фиг. 1. Лазерният сноп се разделя от разцепител на обектен сноп и опорен сноп. Обектният сноп минава през изследвания обект и преминалата през него светлина се детектира от камера. Опорният лъч след отразяването му от фазо-

отместващо огледало минава през обектив и микробленда и осветява камерата. Обективът и опорната микробленди създават сферична опорна вълна. Така получените обектна и опорна вълни имат еднакъв сферичен фронт и интерферират върху фоточувствителния сензор на камерата.



Фиг. 1.

### Математически модел

Когато разстоянието между обекта и регистриращата повърхност е много по-голямо от размерите на обекта, измереният интензитет може да бъде апроксимиран с Фурие образ на обектното изображение [3,4,5]. В този случай пълният интензитет, резултат от интерференцията на обектния и опорния снопове, измерен от камерата, се дава от израза:

$$I = AA^* = |F(O.R_0) + R_1|^2, \quad (1)$$

където  $I$  е интензитетът на всяка точка от изображението на камерата,  $O$  е предавателната функция на обекта,  $R_0$  е разпределението на амплитудата в обектния сноп,  $R_1$  е разпределението на амплитудата в опорния сноп и  $F$  е Фурие трансформацията.

### Алгоритъм за изчисляване

За прилагане четири-стъпков вариант на фазово-стъпковия метод [6] е необходимо да бъдат заснети четири изображения -  $I_0..I_3$  при присъствие на обект и различни фази със стъпка  $\pi/2$  на опорния сноп и четири -  $J_0..J_3$  без обект и със същите стъпки на опорния сноп. Две системи уравнения могат да бъдат съставени.

$$\begin{cases} I_0 = m[1 + a * SIN(\varphi + 0)] \\ I_1 = m[1 + a * SIN(\varphi + 1/2 * \pi)] \\ I_2 = m[1 + a * SIN(\varphi + \pi)] \\ I_3 = m[1 + a * SIN(\varphi + 3/2 * \pi)] \end{cases} \quad (2)$$



$$\begin{cases} J_0 = n[1 + r * \text{SIN}(\psi + 0)] \\ J_1 = n[1 + r * \text{SIN}(\psi + 1/2 * \pi)] \\ J_2 = n[1 + r * \text{SIN}(\psi + \pi)] \\ J_3 = n[1 + r * \text{SIN}(\psi + 3/2 * \pi)] \end{cases} \quad (3)$$

където  $a$  и  $\varphi$  са модулът и фазата на спектъра с обект и  $r$  и  $\psi$  – без него.

$$\begin{aligned} |A| = a &= \sqrt{(I_0 - I_2)^2 + (I_1 - I_3)^2} \\ \arg(A) = \varphi &= \arctan \frac{I_0 - I_2}{I_1 - I_3} \end{aligned} \quad (4)$$

$$\begin{aligned} |R| = r &= \sqrt{(J_0 - J_2)^2 + (J_1 - J_3)^2} \\ \arg(R) = \psi &= \arctan \frac{J_0 - J_2}{J_1 - J_3} \end{aligned} \quad (5)$$

Като използваме обратно Фурие преобразуване получаваме предавателната функция на обекта:

$$O = \frac{F^{-1}(A)}{F^{-1}(R)}, \quad (6)$$

където  $O$  е матрица от комплексни числа, модулът на които е затихването, а аргументът - фазовото отместване за всяка точка от измервания обект.

### Грешки и апроксимации

За да се приложи теорията за Фурие преобразуването, е необходимо да са изпълнени условията измерването да е в далечната (Фраунhoferовата) зона. За първа стъпка при реализацията считаме за важно да приложим успешно алгоритъм за измерване на малки обекти. Пространствените честоти при тях са разположени в по-малък пространствен ъгъл, отколкото при микрообектите. Това ще даде по-добра възможност за сравнение на експеримент и теория. При всички случаи методът ще се прилага за 2D обект. Подходът за снимане на изображения със статична камера дава недостатъчно информация за 3D решение. Също така разсейването и отражението на обекта не се вземат под внимание. Измерването не е чувствително към позицията на камерата. При тази схема наклоняването на камерата е невъзможно поради твърдата връзка към разцепителя. Завъртането и довежда до „завъртане“ на решението, а отместването – до фазова грешка. Калибрирането на камерата чрез подходящ алгоритъм за намиране на центъра намалява влиянието на тези фактори. Както се вижда от фиг. 1 за тази оптична схема, максималният пространствен ъгъл е ограничен от геометрията на куба (разцепителя) преди камерата. Това води до ограничение на пространственото разрешение до около 1.5  $\mu\text{m}$ .

### Програмно осигуряване

За извършване на симулацията е разработена програма, състояща се от две части. Първата генерира изображенията, които биха се видели от камерата. Могат да бъдат зададени различни стойности на параметрите на апертурата на обектния и опорния сноп. Две изображения, представляващи реалната и имагинерната част на предавателната функция на обекта, се генерират и зареждат за обработване. Трябва да бъдат зададени геометрични характеристики като координатите на микроблендите, размерите на сензора на камерата, разстоянието между микроблендите и камерата. Фазово-отместващото устройство е симулирано, като се променя стойността на фазата на опорния сноп. Може да се използва,

който и да е от два вида на изчисляване на интензитета – с бързо Фурие преобразуване или обикновено интегриране чрез сумиране. При високи пространствени честоти бързо Фурие преобразуване не може да се използва, защото условията за това са нарушени. Обикновеното интегриране дава възможност да се симулира използването на няколко камери. Като краен резултат от първата част на програмата се получават осем изображения.

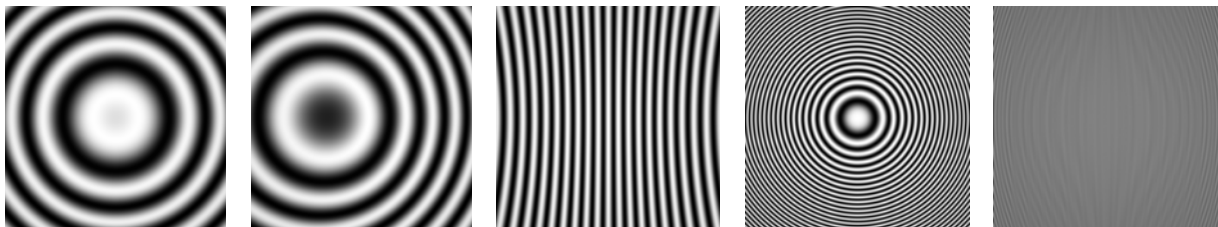
Втората част от програмата извършва обработката на данни от осемте симулирани или измерени изображения. Тя е реализация на уравненията от (2) до (6). При обработването на изображенията е симулирано снимането с различно разрешение на АЦП на камерата. Може да бъде приложен изглаждащ филтър с различни параметри преди обработките. Резултатът от тази част е изчислена предавателна функция на обекта.

Програмата е написана на Microsoft Visual C за WINDOWS.

### Позиция на микроблендите

В идеалния случай на настройване на интерферометъра камерата ще показва плоска картина – с един и същи интензитет по цялата площ. На практика това е много трудно да се постигне, защото при интерференцията на две сферични вълни е необходимо центровете на излъчване (двете бленди) да се съвместят по трите координати. Отместване на блендата от идеалното място по направление на лъча ( $Z$  координата) с  $10\ \mu\text{m}$  при тази геометрия на експерименталната постановка води до поява на интерференчни пръстени с ширина в края на изображението от порядъка на 10 до 30 точки.

Това е максимално допустимата ширина с оглед филтриране на изображението и необходимото понякога възстановяване на непрекъснатостта на фазата. Отместването по другите две оси води до появата на двойно по-тесни хоризонтални или вертикални ивици. На фиг. 2 са показани интерференчните картини, които се получават при различни отмествания по осите – успоредно и перпендулярно на посоката на разпространението на лъча.



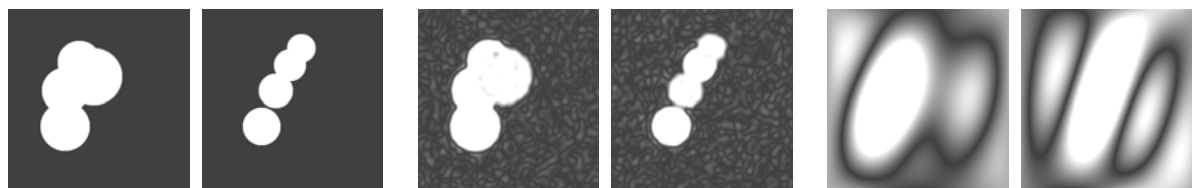
Фиг.2 Интерференчни пръстени при различно отместване по осите  $Z$  и  $X$ .

a)  $X=0, Z=14\ \mu\text{m}$ ;    b)  $X=2\ \mu\text{m}, Z=14\ \mu\text{m}$ ;    c)  $X=14\ \mu\text{m}, Z=0$ ;    d)  $X=0, Z=100\ \mu\text{m}$ ;    e)  $X=100\ \mu\text{m}, Z=0$ ;

### Разрядност на АЦП

При измерване със СЗП камери обикновено се използват от 4 до 14 разрядни АЦП. Охлаждането на фотоелемента на камерата довежда до намаляване на утечните токове, повишава максималното време за експозиция и увеличава съотношението сигнал / шум. За да се оценят изискванията към АЦП на камерата, е симулирано измерване на обект с комплексна предавателна функция. Снимките показват реалната и имагинерната и част.

На фиг. 3 са показани изходната стойност на параметрите на обекта, както и стойностите при симулацията на две измервания с 6-битово АЦП и еднобитово АЦП. За по-голяма точност естествено е необходимо АЦП с повече разряди. Симулацията показва, че точността, получена при използването на 6-битово АЦП, е достатъчна за визуализиране на измерения обект. Грешката в случая е 2%. Дори използването на еднобитово АЦП води до приемливи резултати от измерването. Това е така, защото в измерването участват голям брой точки от камерата.



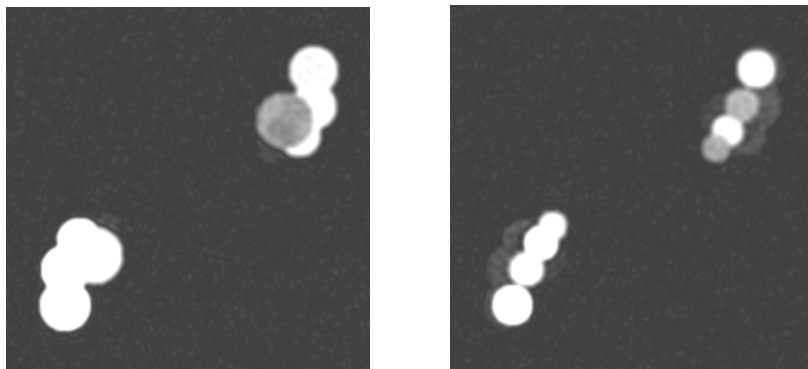
Фиг. 3 Реална и имагинерна част на:

а) зададен обект;

б) симулирано измерване  
с 6-битово АЦП;в) симулирано измерване  
с 1-битово АЦП

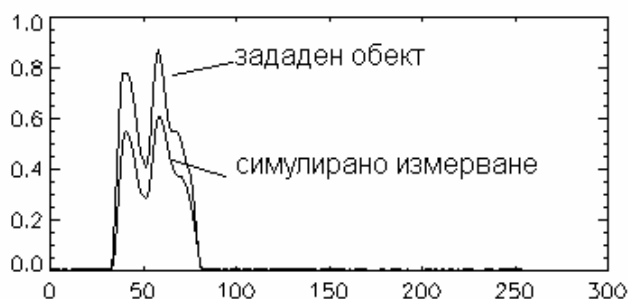
### Грешка при фазовото отместване

Фазово-отместващото устройство може да бъде калибрирано с използване на спомагателен интерферометър [7], който не е показан на схемата и обратна връзка към компютъра от смет сигнал от допълнителен фото сензор. Обикновено поради чувствителността на интерферометричната постановка околните паразитни вибрации се предават на огледалото и отместването на опорния спрямо обектния лъч не е точно зададената стойност. Тези вибрации са неизбежни и е необходимо да се оцени тяхното влияние. Симулирани са грешки в отместванията с Гаусово разпределение с различна амплитуда. При повишаване нивото на грешката стойността на реалната и имагинерната част на обекта се намаляват и се появява паразитно изображение, симетрично относно центъра на картината. Резултатите от едно такова измерване при грешка на фазите 1 радиан е показано на фиг. 4.



Фиг. 4 Реална и имагинерна част при грешка във фазовото отместване

На фиг. 5 е показано едно сечение на стойността на амплитудата, минаващо през симулирания обект.



Фиг. 5 Амплитуда на едно сечение при грешка във фазовото отместване

---

**Заклучение**

---

Естеството на измерването налага изработването на стенд-модел с висока точност и възможност за фини регулировки на позицията на някои оптични елементи. Регулировката трябва да е, както на оптичните елементи един спрямо друг (обектив и бленда), така и на двете оптични групи (обектна и опорна) една спрямо друга. Точността на регулировка трябва да е по-добра от 10 мкм. Настройката на оптичната система, чистотата на оптичните елементи, просветляването на CCD-сензора, отражението от предната и задната му повърхности довеждат до появата на паразитни интерференчни пръстени. Те са постоянни като изображения и тяхното влияние се елиминира, като се използва опорно измерване. Модулацията, която се получава вследствие непрекъснатото им присъствие в изображението, намалява съотношението сигнал–шум в затъмнените участъци. Мощността на лазера и чувствителността на камерата са важни, защото при висока плътност на енергията е възможна повреда на измервания обект. Понижаването на интензитета на светлината в дадена точка с 50 % е равносилно на намаляването на разредността на АЦП с един бит. Ето защо увеличението на разредността на АЦП на камерата би повишило чистотата на изчисленото изображение и би намалило влиянието на някои от грешките от настройката на оптичната система.

---

**Литература**

---

1. Baltes H. P., Inverse Source Problems in Optics, Springer-Verlag Berlin-New York, 1978.
2. Tiziani H.J., Optical methods for precision measurements, Optical and Quantum Electronics 21 (1989) 253-282.
3. Cowley J. M., Diffraction Physics, North-Holland Publishing Company, New York, 1975.
4. Rastogi P., Sharma A., Systematic approach to image formation in digital holography. Opt. Eng. 42(5) 1208-1214 (May 2003).
5. G. Stoilov, N. Mechkarov, P. Sharlandjiev, „Information Modelling of Two-Dimensional Optical Parameters Measurement”, Information Theories & Applications, vol. 11, (2004)
6. Wayant J.C., “Interferometric optical metrology: basic systems and principals”, Laser Focus, 65-71, may 1982
7. V. Sainov, G. Stoilov, D. Tonchev, T. Dragostinov, A. Dimitrova, “An Optoelectronic Feedback System For Measuring The Shape Of Large- Format Objects”, San Diego, Proc. SPIE, v.2545 (1995)

---

**Информация за автора**

---

**Georgi Stoilov** – CLOSPI, research scientist, CLOSPI- BAS, Sofia-1113, Acad. G. Bontchev Str.- 101, P.O. Box 95, [gstoilov@optics.bas.bg](mailto:gstoilov@optics.bas.bg)

---

---

## Informatics in Education and Cultural Heritage

---

---

### **DIGITISATION OF CULTURAL HERITAGE: BETWEEN EU PRIORITIES AND BULGARIAN REALITIES<sup>1</sup>**

**Milena Dobрева, Nikola Ikonov**

**Abstract:** *The paper presents the Bulgarian setting in digital preservation of and access to cultural and scientific heritage. It mentions key Bulgarian institutions, which take or should take part in digitisation endeavours. It also presents examples of building and adapting specialised tools in the field, and more specifically SPWC, ACT and XEditMan.*

**Keywords:** *digital preservation of and access to cultural and scientific heritage, legislative issues, SPWC, ACT, XEditMan*

---

#### **Introduction**

---

Before the current period of economic transition, until 1990, Bulgaria was the Eastern-European country with highest expertise in information technologies within the CMEA (Council for Mutual Economic Assistance). During this time, digitisation of cultural and scientific heritage was not a separate field of work – but domains such as library automation systems, databases of cultural heritage objects; corpora of ancient and mediaeval texts had been already developed for several decades. Such work was also done in Bulgaria but not on a large-scale systematic basis.

In the years after 1990, Bulgaria has been undergoing a period of structural changes and economic transition. The acquired technological excellence in information and communication technologies (ICT) had been transferred from huge institutions to small and medium-sized enterprises functioning in a highly competitive environment. Digitisation activities, which require large investments and are not bringing quick profit, have not become attractive for the companies from the ICT sector.

Additionally, culture, education and science sectors have been suffering from inadequate funding during the transition period (the share of gross national product spent for science for example in the last years is about 0,29% which is 10 times less than in the EC). This general setting was not favourable for the establishing of national and institutional digitisation programmes.

At the same time, Bulgarian collections house over 12,500 manuscripts of Slavonic, Greek, Latin, Ottoman Turkish and other origin. Another key example is the epigraphic inscriptions from the Antiquity period, which form the third largest collection in the world following those of Italy and Greece. Precious monuments of immovable heritage, nine objects in the UNESCO World Heritage List, numerous archæological findings, Old Bulgarian runic inscriptions—all these materials are of interest not only for the local community, but also for the wider European community of which Bulgaria is a cultural part and indeed for all of mankind globally by virtue of the shared

---

<sup>1</sup> The part of this article presenting institutional involvement is an extended and updated version of the respective part of a report on Bulgaria prepared by Milena Dobрева for a MINERVAPlus Global Report (**Coordinating digitisation in Europe**. Progress report of the National Representatives Group: coordination mechanisms for digitisation policies and programmes 2004, forthcoming).

meaning of the culture of the Other. Yet, electronic information on these resources is still hardly accessible in its fullness not only to foreign experts, but also to regional and local specialists.

In this paper we will present the key types of institutions playing different roles in digitization processes with a brief comment on their actual involvement in digitisation.

---

### **Bulgarian Institutions and Their Level of Involvement in Digitisation Activities**

---

Six types of organisations are potentially interested in digitisation of cultural heritage: government bodies, repositories, research and/or educational institutions, companies, and foundations.

**Government bodies** are entrusted with the supervision of such activities.

Three institutions should play the key role for establishing digitisation policy in Bulgaria but neither one is currently working in this direction:

1. *The Ministry of Culture and Tourism*.<sup>1</sup> The last structural change was done very recently, in the end of January 2005 when Tourism was added to the activities of the ministry in recognition of the fact that cultural tourism will be one of the basic specialisation sectors for the Bulgarian economy in the next years.
2. *ICT Development Agency at the Ministry of Transport and Communications*<sup>2</sup>, Republic of Bulgaria. The agency was created in 1995 and currently is the body responsible for the development of the information and communication technologies in Bulgaria. In the last few years it provided funding for projects aimed at presentation of cultural heritage in electronic form. One example is the first XML repository of catalogue descriptions of Old Bulgarian manuscripts preserved in Bulgaria – a project carried out by the Institute of Mathematics and Informatics and funded by the Agency in 2004. However, a coherent strategy has not been created and respectively followed.
3. *Ministry of Education and Science*.<sup>3</sup> Digitisation is not a technical activity; it develops rapidly and involves new research results.

**Repositories** (libraries, archives and museums), which seem the most natural initiators of digitisation projects because of the close relationship between digitisation and preservation, are currently in the position of observers due to lack of funding on the one hand, and copyright issues for digital collections, on the other hand. There are about 7000 public, university, scientific, specialised libraries and information centres in the country. As most important institutions in this group we should mention:

1. *General Department of Archives at the Council of Ministers of Republic of Bulgaria*.<sup>4</sup> The General Department of Archives initiated pilot work in digitisation of archival documents with the publication of the documentary CD compendium “The Independence of Bulgaria and the Bulgarian Army” containing materials from the Central Military Archive in Veliko Turnovo in 2003. The vision on digitisation activities of the State Archives was presented recently [Markov 2004].
2. *The National Library “Saint Cyril and Saint Methodius”*<sup>5</sup> plays a leading role in the process of expert decision-making related to measures of digital cataloguing and publishing of mediæval manuscript heritage and early printed books. Its prescriptions in these fields are adopted in other libraries in the country, which have such collections. The National Library is also the basic driving force for digital cataloguing of modern books. Although the library experts have quite extensive experience in following the current practices, real digitisation work has not been planned [Moussakova, Dipchikova 2004].
3. *The National Museum of History*<sup>6</sup> does not seem to be currently involved in any digitisation-related work.

---

<sup>1</sup> <http://www.culture.government.bg/> date of last visit 25.5.2005.

<sup>2</sup> <http://www.ict.bg/>, date of last visit 25.5.2005.

<sup>3</sup> <http://www.minedu.government.bg/>, date of last visit 26.5.2005.

<sup>4</sup> [http://www.archives.government.bg/index\\_en.html](http://www.archives.government.bg/index_en.html), date of last visit 25.5.2005.

<sup>5</sup> <http://www.nationallibrary.bg/>, page does not open on 25.5.2005.

<sup>6</sup> <http://www.historymuseum.org/>, date of last visit 25.5.2005.

Research and/or educational institutions are the most active initiators of small-scale digitisation projects in Bulgaria. They usually do not have the funds and resources for running mass digitisation projects, but are the most active promoters of this field of work.

1. *The Institute of Mathematics and Informatics*<sup>1</sup> of the Bulgarian Academy of Sciences (IMI) plays the leading role in this direction. Digitisation of Scientific Heritage department<sup>2</sup> was established in IMI in 2004. The institute took part in projects related to digitisation of mathematical heritage; cataloguing and electronic publishing of mediæval Slavonic manuscripts. In addition, IMI organised in the last years three summer schools and four specialised workshops related to digitisation of cultural and scientific heritage which were targeted at Central European countries' participants and have regional impact.

The Institute produced the most extensive XML catalogue (over 800 catalogue records) of Old Bulgarian manuscripts stored in Bulgaria [Pavlov 2004] in cooperation with specialists from the Faculty of Mathematics and Informatics of the Sofia University "Kliment Ohridski" and the National Library "St Cyril and St Methodius" IMI is the coordinator of the international project Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria (KT-DigiCult-BG), supported by the Marie Curie programme, Framework Programme 6 of the European Commission which is implemented in 2004-2008.

The Institute is a partner in the COMTOOCI project supported by eCulture program of the EC, which is coordinated by the Institute for computational linguistics in Pisa, Italy. In this project its role is to support the localisation and local implementation of a specialised software for philological and librarian work in the cultural institutions which was developed by the Italian institute.

IMI also works on presentation of folklore archives in digital form in cooperation with the Institute for folklore of the Bulgarian Academy of Sciences.

2. *The Institute for Bulgarian Language*<sup>3</sup> (IBL) works on digital preservation and use of audio archives containing live recordings presenting various Bulgarian dialects. These records originally were collected in the 50s and 60s in the 20c, and their conversion in electronic form was absolutely necessary since the original tapes started to deteriorate.
3. Amongst educational institutions we should mention The State Library Institute,<sup>4</sup> which recently opened a specialized programme *Information funds of the cultural and scientific heritage*. Sofia University offers a general programme on Library and information activities<sup>5</sup>.

**Companies** are interested in presenting sections of cultural heritage to the world, which they believe will be easily realised on the market. Today it is rather difficult to establish customer interest. The Bulgarian market for such products is unsatisfactory. This is why their main market is abroad. As an example of a company, which specializes in digitisation services, we could mention BalkanData<sup>6</sup> - a US-owned company based in Bulgaria. This combination seeks to offer the winning combination of the local technological and intellectual excellence and the low labour costs in the country.

**Non-governmental institutions (NGOs).** One active organisation in the library field is The Union of Librarians and Information Services Officers (ULISO).<sup>7</sup> It produced in 1997 the National Program for the preservation of Library Collections.

**Funding bodies (foundations)** rarely support projects undertaken in the field of digitisation. In addition, the scale of their support cannot meet the real costs of serious digitisation projects. In the last years the tendency is that such bodies are supporting basically dissemination activities (workshops, conferences, trainings).

<sup>1</sup> [www.math.bas.bg](http://www.math.bas.bg), date of last visit 25.5.2005.

<sup>2</sup> <http://www.math.bas.bg/digi/indexbg.html>, date of last visit 25.5.2005.

<sup>3</sup> <http://www.ibl.bas.bg>, date of last visit 25.5.2005.

<sup>4</sup> <http://www.svubit.org/>, date of last visit 25.5.2005.

<sup>5</sup> <http://forum.uni-sofia.bg/filo/display.php?page=bibliotekoznanie>, date of last visit 25.5.2005.

<sup>6</sup> <http://www.balkandata.net/>, date of last visit 25.5.2005.

<sup>7</sup> <http://www.lib.bg/act.htm>, date of last visit 25.5.2005.

---

### Legislative Issues

---

The main cultural and scientific heritage collections in Bulgaria belong to the State and their maintenance is totally dependent on the State budget. One would expect that the development of a national policy for digitisation would be an easy task when most collections of the cultural heritage are State-owned. Unfortunately, most of the legislation in the cultural sphere does not cover any digitisation aspects. A brief presentation of key legal acts covering issues, which could be approached also in digitisation programmes follow.

The Law for Protection and Development of the Culture<sup>1</sup> (in force since 1 January 2001) defines the basic principles and functions of the national cultural policy and the cultural institutions. However, digitisation is not mentioned amongst the issues that are covered in it.

The Deposit Law<sup>2</sup> (last version in force as of 1 January 2001) addresses works on digital media (electronic documents). According to it, works published on digital media should be presented in three copies to the National Library within two weeks after the publication. The National Library stores these materials as physical copies, and is not seen as a body, which would include the electronic publications into a digital library.

The Regulation for Rendering and Saving Movable Cultural Monuments<sup>3</sup> addresses the matters of finding, collecting, and preserving of movable cultural heritage monuments and making scientific descriptions related to them. Its application is mandatory for all museums, art galleries, museum collections as well as individuals. According to Article 62, the basic form of record and scientific description is the inventory book. The detail and accuracy of records is the responsibility of the directors of the collections. The scientific descriptions of immovable objects are presented as "Scientific passports" of the objects (Article 79). This regulation is in force since 1 January 1974. Understandably, electronic records and links between documentation of various collections were not planned in that time, but changes, which would take into account the current state of technology, have not been made.

The Regulation N 26 of 10.04.1996 of the Development, Use and Management of an Automated Information System "An Archæological map of Bulgaria"<sup>4</sup> seems to be the only legislative act in Bulgaria which treats a matter of digital presentation and storage of data related to the cultural heritage. It addresses the development of a specialized information system. The feeding of the database is the responsibility of the Institute of Archæology of the Bulgarian Academy of Sciences and the National Institute for Cultural Monuments based on primary data supplied from specialists who worked *in situ*. Information can be obtained from this automated system only on the basis of a written request for a service fee. The collection of data and their use were adequate for the state of the technologies in 1996; now this is outdated but changes to adapt the collected data and to provide access via the Internet have not been done.

The Tariff of rates collected by State Cultural Institutions for Services and Provision of Documents and Copies<sup>5</sup>, date of last update 5 January 2001 does not include any fees related to digital images despite the recency of the update.

---

### EU cooperation and Current EU priorities

---

Bulgarian institutions are active in searching for international cooperation possibilities. Within the trend of Digital culture (Access to and preservation of cultural heritage) in FP6, we can mention the following projects where Bulgarian institutions participate as members:

- CALIMERA (participant ULISO)
- EPOCH (participant New Bulgarian University)

---

<sup>1</sup> <http://www.culture.government.bg/docdetail.html?id=16>, in Bulgarian, date of last visit 25.5.2005.

<sup>2</sup> <http://www.culture.government.bg/docdetail.html?id=66>, in Bulgarian, date of last visit 25.5.2005.

<sup>3</sup> <http://www.culture.government.bg/docdetail.html?id=49>, in Bulgarian, date of last visit 25.5.2005.

<sup>4</sup> <http://www.culture.government.bg/docdetail.html?id=48>, in Bulgarian, date of last visit 25.5.2005.

<sup>5</sup> <http://www.culture.government.bg/docdetail.html?id=38>, in Bulgarian, date of last visit 25.5.2005.



- MINERVAPLUS (participant – IMI-BAS as an associated member)
- PRESTOSPACE (participant Sirma AI Ltd)
- KT-DigiCult-BG is a project coordinated by IMI-BAS.

IMI-BAS was an initiator of the creation of the South-Eastern European Network for Digitisation of Scientific and Cultural Heritage<sup>1</sup>, constituted with the signing of the Borovets declaration of 17 September 2003.

The current priorities under IST 2.5.10 (Access to and preservation of cultural and scientific resources) are targeted to:

- Enriched conceptual representations
  - Advanced access methods
  - Long-term preservation
- The presentation of current Bulgarian setting in the previous sections shows that some Bulgarian institutions are trying to be in line with current developments.

---

### Development of Local Tools vs Adaptation of Existing Tools

---

In the digitisation work one crucial matter is what tools will be applied for the practical work. In the last years IMI gained experience in two approaches: *developing local tools* for support of specific task and *localisation of existing platforms* to the Bulgarian environment.

As a **home-made tool** we could mention XEditMan [Pavlov 2004]. This is a tool, which combines an editor and visualisation component for preparing and studying manuscript description of mediaeval manuscripts. Its interface is in Bulgarian and follows the local practices in cataloguing work. The descriptions of manuscripts are produced in XML format following the TEI P4 guidelines.

The experience with development and use of this tool is very positive, since it supports the performance of a specific task and facilitates the preparation of large amount of data in digital form.

As an example of **localised tool**, we could mention **SPWC**, Software Platform for archivist, librarian and philological Work in Cultural Institution. The platform offers a set of tools for document management in cultural institution including digitizing, cataloguing and transcription of primary sources. ILC are active in building specialized workstations for philological work for decades [Bozzi, Corradini 2004] and constantly improve the capabilities and the spread of use of their specialized tools. IMI worked on localisation of the software (translation of the user interface and documentation in Bulgarian), identification of experimental materials (in the case of Bulgaria this are local DTDs – manuscript and archival records), and training and dissemination activities. All abovementioned endeavours were in the frame of the project COMTOOCI (COMputational TOOlS for the librarian and philological work in Cultural Institution). The project, supported by the CULTURE 2000 program and coordinated by the Institute for Computational Linguistics (ILC) – Pisa, Italy started in September 2004. In the forthcoming months a pilot installation of SPWC in the General Department of Archives will be done.

At the same time, we are studying the possible application of **ACT** [Ribarov 2004]. This software combines the presentation of manuscript images and annotated texts. Because of the high level of variety in mediaeval Slavonic manuscripts, the author chose an approach where previous human annotation of wordforms is used in subsequent annotation activities.

We present a comparison of the features of **ACT** and **SPWC** in Table 1.

---

<sup>1</sup> <http://www.ncd.matf.bg.ac.yu/?page=news&lang=en&file=declaration.htm>, date of last visit 25.5.2005.

FEATURES	ACT	SPWC
Image analysis module		✓
Image representation module	✓	✓
Cataloguing module		✓
Representation of texts	✓	✓
Representation of variants	✓	✓
Multilingual support	✓	✓
Interface in different languages		✓
Annotation of various levels, up to morphology	✓	
Morphological annotation supported on 'learning by example' basis	✓	

---

## Conclusion

Under the described lack of national policy, the various institutions in the cultural and scientific heritage sector have the freedom to design their own policies. Unfortunately, this is combined with lack of methodological, financial, technological and human resources support. On this setting, the Digitisation of Scientific Heritage Department at IMI has as a core part of its mission to contribute to the improvement of human resources qualification and support memory institutions through joint activities, which would lead to a difference in the future.

---

## Bibliography

- [Bozzi, Corradini 2004] Bozzi A., Corradini M.S. (2004). Aspects and methods of computer-aided textual criticism. In Digital Technology and Philological Disciplines. *Linguistica Computazionale XX-XXI*, (2004). Pisa-Roma, IEPI. 49-66.
- [Markov 2004] Markov N., National Archives. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August—3 September 2004). Vol. 11 (2004), No. 3, pp. 282-283.
- [Moussakova, Dipchikova 2004] Moussakova E, A. Dipchikova, The Role of the National Library in Preserving National Written Heritage. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August—3 September 2004). Vol. 11 (2004), No. 3, pp. 284-287.
- [Pavlov 2004] Pavlov P., XML Presentation of Catalogue Data on Medixval Slavonic Manuscripts: Experience and Perspectives, In: Proceedings of the 33rd Conference of the Union of Bulgarian Mathematicians, Borovets, 1–4 April 2004, pp.236–240.
- [Ribarov 2004] K. Ribarov *et al.*, "We present the ACT Tool". In: Scripta & e-Scripta, Volume 2, Institute of Literature, Bulgarian Academy of Sciences, Sofia. 2004.

---

## Authors' Information

**Milena Dobreva** – Chair of the Department *Digitisation of Scientific Heritage*, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev St., bl. 8, Sofia-1113, Bulgaria, e-mail: [dobрева@math.bas.bg](mailto:dobрева@math.bas.bg)

**Nikola Ikonov** – Chair of Laboratory on Phonetics and Speech Communication, Institute for Bulgarian Language, BAS, Shipchenski prohod 52, Sofia-1113, Bulgaria, e-mail: [nikonomov@ibl.bas.bg](mailto:nikonomov@ibl.bas.bg).

---

## PROGRAMMING PARADIGMS IN COMPUTER SCIENCE EDUCATION

Elena I. Bolshakova

**Abstract:** *Main styles, or paradigms of programming – imperative, functional, logic, and object-oriented – are shortly described and compared, and corresponding programming techniques are outlined. Programming languages are classified in accordance with the main style and techniques supported. It is argued that profound education in computer science should include learning base programming techniques of all main programming paradigms.*

**Keywords:** *programming styles, paradigms of programming, programming techniques, integration of programming techniques, learning programming paradigms*

---

### Introduction

Several main **styles** (or **paradigms**, or models) **of programming** – imperative, functional, logic and object-oriented ones – were developed during more than forty-year history of programming. Each of them is based on specific algorithmic abstractions of data, operations, and control and presents a specific mode of thinking about program and its execution. Various **programming techniques** (including data structures and control mechanisms) were elaborated rather independently within each style, thereby forming different scopes of their applicability. For instance, the object-oriented style and corresponding techniques are suitable for creating programs with complicated data and interface, while the logic style is convenient to program logic inference.

Though modern programming languages [Finkel, 1996] usually include programming techniques from different styles, they may be classified according to the main style and techniques supported (e.g., programming language Lisp is a functional language while it includes some imperative programming constructs).

Nowadays, for implementation of large programming project, techniques from different paradigms are required, mainly because of complexity and heterogeneity of problems under solution. Some of them are problems of complex symbolic data processing, for which programming techniques of functional and logic languages (e.g., Lisp [Steele, 1990] or Prolog [Clocksin, 1984]) are adequate. The other problems can be easily resolved by means of popular imperative object-oriented languages, such as C++ [Stroustrup, 1997].

Below we explain our point that acquirement of programming techniques of all main paradigms belong to background knowledge in the field of computer science. Accordingly, learning of modern programming languages should be complemented and deepened by learning of programming paradigms and their base techniques.

---

### Programming Paradigms

The **imperative** (procedural) programming paradigm is the oldest and the most traditional one. It has grown from machine and assembler languages, whose main features reflect the John von Neuman's principles of computer architecture. An imperative program consists of explicit commands (instructions) and calls of procedures (subroutines) to be consequently executed; they carry out operations on data and modify the values of program variables (by means of assignment statements), as well as external environment. Within this paradigm variables are considered as containers for data similar to memory cells of computer memory.

The **functional** paradigm is in fact an old style too, since it has arisen from evaluation of algebraic formulae, and its elements were used in first imperative algorithmic languages such as Fortran. Pure functional program is a collection of mutually related (and possibly recursive) functions. Each function is an expression for computing a value and is defined as a composition of standard (built-in) functions. Execution of functional program is simply application of all functions to their arguments and thereby computation of their values.

Within the **logic** paradigm, program is thought of as a set of logic formulae: axioms (facts and rules) describing properties of certain objects, and a theorem to be proved. Program execution is a process of logic proving (inference) of the theorem through constructing the objects with the described properties.

The essential difference between these three paradigms concerns not only the concept of program and its execution, but also the concept of program variable. In contrast with imperative programs, there are neither explicit assignment statements nor side effects in pure functional and logic programs. Variables in such a program are similar to those in mathematics: they denote actual values of function arguments or denote objects constructed during the inference. This peculiarity explains why functional and logic paradigms are considered as non-traditional.

Within the **object-oriented** paradigm, a program describes the structure and behavior of so called objects and classes of objects. An object encapsulates passive data and active operations on these data: it has a storage fixing its state (structure) and a set of methods (operations on the storage) describing behavior of the object. Classes represent sets of objects with the same structure and the same behavior. Generally, descriptions of classes compose an inheritance hierarchy including polymorphism of operations. Execution of an object-oriented program is regarded as exchange of messages between objects, modifying their states.

**Table 1.** Features of the main programming paradigms

<b>Paradigm</b>	<b>Key concept</b>	<b>Program</b>	<b>Program execution</b>	<b>Result</b>
Imperative	Command (instruction)	Sequence of commands	Execution of commands	Final state of computer memory
Functional	Function	Collection of functions	Evaluation of functions	Value of the main function
Logic	Predicate	Logic formulas: axioms and a theorem	Logic proving of the theorem	Failure or Success of proving
Object-oriented	Object	Collection of classes of objects	Exchange of messages between the objects	Final state of the objects' states

The object-oriented paradigm is the most abstract, as it's basic ideas can be easily combined with the principles and programming techniques of the other styles. Really, an object method may be interpreted as a procedure or a function, whereas sending of message as a call of procedure or function. Contrarily, traditional imperative paradigm and non-traditional functional and logic ones are poorly integrated because of their essential difference.

Distinguishing features of the main programming paradigms are clarified in Table 1.

### **Programming Languages and Programming Techniques**

Each algorithmic language was initially evolved within a particular paradigm, but later it usually accumulates elements of programming techniques from the other styles and languages (genesis of languages and relations between them are shown in Fig.1). Hence, as a rule, most languages include a kernel comprising programming techniques of one paradigm and also some techniques from the other paradigms. We can classify languages according to paradigms of their kernels. The following is a classification of several famous languages against the main paradigms:

- Imperative paradigm: Algol, Pascal, C, Ada;
- Functional paradigm: Lisp, Refal, Planner, Scheme;
- Logic paradigm: Prolog;
- Object-oriented paradigm: Smalltalk, Eiffel.

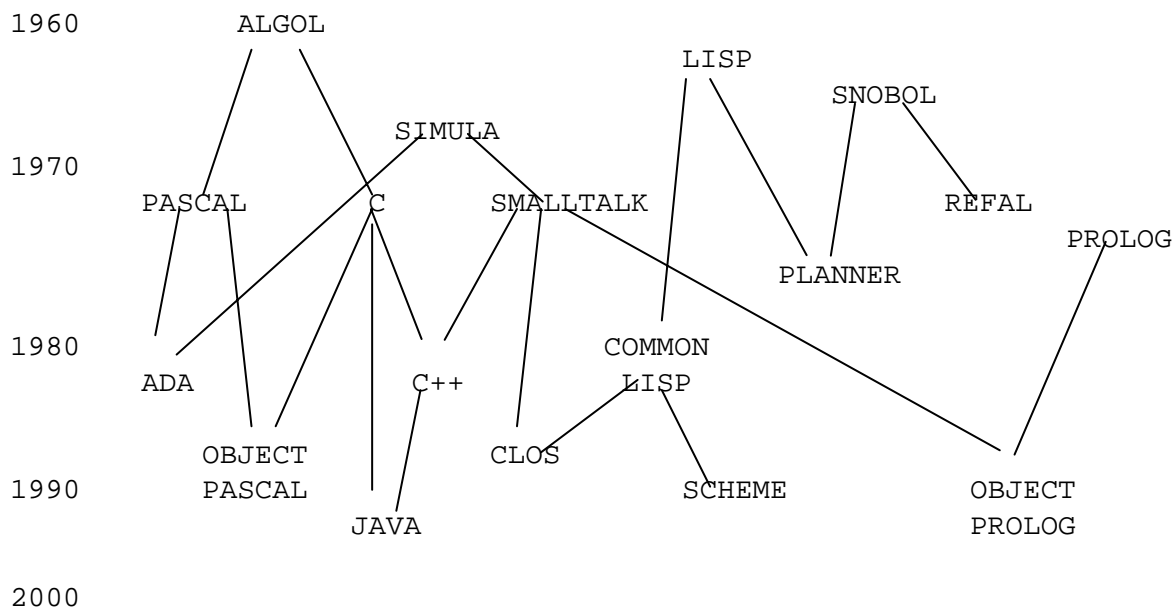
We could notice that Smalltalk [Goldberg, 1982], the first object-oriented language, is not popular because of complexity of its syntax and dynamic semantics. But its basic object ideas (abstraction of object's state and behavior, encapsulation and inheritance of state and behavior, polymorphism of operations) are easily integrated with the principles of programming languages of the other styles. For this reason, the object-oriented paradigm became widespread as soon as it was combined with traditional imperative paradigm. To be more precise, it

became widespread when it was embedded into the popular imperative languages C and Pascal, thereby giving imperative object-oriented languages C++ and Object Pascal.

Analogous integration of object-oriented principles with programming techniques of the other paradigms has led to object-oriented variants of non-traditional languages. For example, the language Clojure is an object-oriented Lisp developed on the base of Common Lisp [Steele, 1990], the popular version of Lisp. Modern programming languages, which are combinations of two paradigms, are:

- Imperative + Object-oriented paradigms: C++, Object Pascal, Ada-95, Java;
- Functional + Objects-oriented paradigms: Clojure;
- Logic + Object-oriented paradigms: Object Prolog.

Programming techniques elaborated within the traditional imperative paradigm and imperative languages, are well known [Finkel, 1996]: control structures include cyclic and conditional statements, procedures and functions, whereas data structures comprise scalar and compound types – arrays, strings, files, records, etc. Programming techniques of imperative object-oriented languages also includes object types and corresponding techniques, such as virtual procedures and functions.



**Fig. 1.** Genealogy of programming languages

Programming languages based on non-traditional paradigms provide rather different data types and control structures [Field, 1988], they also differ from traditional languages in the mode of execution: interpretation instead of compilation applied for imperative and imperative object-oriented languages.

Unlike imperative languages, logic and functional languages are usually recursive and interpretive, and most of them are oriented towards symbolic processing. Besides recursion, programming techniques developed within these languages include:

- flexible data structures for representing complex symbolic data, such as list (Lisp) or term (Prolog);
- pattern matching facilities (Refal, Prolog) and automatic backtracking (Planner, Prolog);
- functionals, i.e. high order functions (Lisp, Scheme);
- mechanism of partial evaluations (Refal).

Programming techniques elaborated within corresponding programming style and programming languages have its own scope of adequate applications. Functional programming is preferable for symbolic processing, while logic

programming is useful for deductive databases and expert systems, but both of them are not suitable for interactive tasks or event-driving applications. Imperative languages are equally convenient for numeric and symbolic computations, giving up to most of the functional languages and Prolog in the power of symbolic processing techniques. The object paradigm is useful for creating large programs (especially interactive) with complicated behavior and with various types of data.

---

### Integration of Programming Techniques

---

Nowadays, imperative object-oriented languages C++, Java, and Object Pascal supported by a large number of developing tools are the most popular choice for implementation of large programming projects. However, these languages are insufficiently suitable for implementation of large software projects, in which one or several problems often belong to the symbolic processing domain where non-traditional languages, such as Lisp or Prolog, are more adequate. For instance, development of a database with a complex structure (approx. a hundred of various relations) and with a natural language interface (queries written as sentences from a restricted subset of natural language and responses in a convenient NL form) involves the following problems to be resolved:

<b>Problems</b>	<b>Suitable programming languages</b>
Syntactic analysis of NL query	Refal
Semantic analysis of the query	Lisp, Prolog
Processing of the query	Prolog
Elaboration of response	Lisp, Refal
Modern user interface	C++, Object Pascal, Java
DB managing operations	C++

On the right hand of the problems, corresponding adequate programming languages are indicated. Evidently, languages oriented to symbolic processing are preferable for syntactic and semantic analysis of natural language queries, as well as for generation natural language phrases expressing responses. We suppose that semantic analysis of the queries may imply some logic inference, which is available in Prolog.

Thus, in order to facilitate implementation of programming projects an integration of programming techniques from different languages and styles is required. In particular, it seems attractive to enhance the power of popular imperative object-oriented languages with special data structures and control mechanisms from non-traditional languages.

As far as the necessity of integration of various programming techniques arisen long before the appearance of popular object-oriented languages, two simplest ways were proposed for solving the problem. The first way suggests creating in the source language some procedural analogue of the necessary technique from another language. This way is labor-intensive and does not preserve the primary syntax and semantics of built-in techniques.

Another way of integration involves coding each problem in an appropriate programming language and integrating of resulting program modules with the aid of multitask operating system (for example, via initiating its own process for each module). This way is difficult to realize because of closed nature of implementation of non-traditional languages and incompatibility of data structures from different languages (e.g., data structures of Prolog and C++ are incompatible).

However, the first way of integration was recently developed for integrating various programming techniques on the basis of an imperative object-oriented language, such as C++ or Object Pascal. The key idea of the method proposed in [Bolshakova and Stolyarov, 2000] is to design, within the given object-oriented language, special classes of objects and operations modeling necessary data and control structures from another language. The method was successfully applied for building functional techniques of the programming language Lisp [Steele, 1990] into the C++ language [Stroustrup, 1997], resulting in a special library of C++ classes that permits to write Lisp-like code within a C++ program.

---

We should also note that the second way of integration, i.e. direct integration of programming codes written in different languages, now becomes perspective in connection with the development of Microsoft.NET platform, which permits compiling and linking such different codes.

---

### Learning Programming Paradigms

---

Necessity of integrating various programming techniques and languages within the same software project is the claim of modern programming. Therefore, a profound education in the field of computer science should be based on learning programming techniques of different paradigms. This implies learning several different algorithmic languages, as none of languages can comprise all possible techniques from various programming styles. Our point is that courses on modern programming languages included in typical curricula should be complemented by special lectures devoted to programming paradigms and intended to compare their base programming techniques and to explicate distinguishing features of the paradigms. Another option to deepen the knowledge of programming techniques and programming languages is to enrich a general course on programming languages with education material on the main programming paradigms. Since the 80s a similar course is read at Algorithmic Languages Department of Faculty of Computational Mathematics and Cybernetic in Moscow State Lomonossov' University.

The importance of learning programming paradigms is also explained by the fact that we can not know the future of popular modern languages. During the history of programming, many languages became dead, while some other languages have lost their popularity, so some modern languages may have the same destiny. However, the main programming paradigms will be the same, as well as their base programming techniques, and thus their learning is a permanent constituent of education in the field of computer science.

---

### Conclusion

---

We have outlined main programming paradigms, as well as programming techniques and programming languages elaborated within them. Programming techniques of traditional imperative paradigm essentially differ from techniques of nontraditional ones – functional and logic. They have different scopes of applicability, and for this reason necessity to integrate techniques of different paradigms often arises in programming projects. Accordingly, a profound education in computer science implies acquirement of programming techniques of all main paradigms, and usual learning of modern programming languages should be complemented by learning of programming paradigms and their base programming techniques.

---

### Bibliography

---

- [Bolshakova, Stolyarov, 2000] Bolshakova, E., Stolyarov A. Building Functional Techniques into an Object-oriented System. In: Knowledge-Based Software Engineering. T. Hruska and M. Hashimoto (Eds.) Frontiers in Artificial Intelligence and Applications, Vol. 62, IOS Press, 2000, p. 101-106.
- [Clocksin, 1984] Clocksin, W.F., Mellish C.S. Programming in Prolog, 2nd edition. Springer-Verlag, 1984.
- [Goldberg, 1982] Goldberg, A., Robson D. Smalltalk-80 – the language and it's implementation. Addison-Wesley, 1982.
- [Field, 1988] Field, A., Harrison P. Functional Programming. Addison-Wesley, 1988.
- [Finkel, 1996] Finkel, R.A. Advanced Programming Language Design. Addison-Wesley Publ. Comp., 1996.
- [Steele, 1990] Steele, G. L. Common Lisp – the Language, 2nd edit. Digital Press, 1990.
- [Stroustrup, 1997] Stroustrup, B. The C++ Programming Language, 3rd edition. Addison-Wesley. 1997.

---

### Author's Information

---

**Elena I. Bolshakova** – Moscow State Lomonossov University, Faculty of Computational Mathematics and Cybernetic, Algorithmic Language Department, Docent; Leninskie Gory, Moscow State University, VMK, Moscow 119899, Russia; e-mail: [bolsh@cs.msu.su](mailto:bolsh@cs.msu.su)

## TUTORING METHODS IN DISTANCE COURSE «WEB-DESIGN TECHNOLOGIES»

Valeriy Bykov, Yuriy Zhook, Ganna Molodykh

**Abstract:** *Methods of Organizational Forms of students' work at the example of distance course "Web-Design Technologies" are described, learning results analysis being carried by the author are suggested. Some concrete recommendations about Tutor's activities in distance learning process are proposed to improve the quality and efficiency of distance learning.*

**Keywords:** *Distance Course "Web-Design Technologies", methods and methodology of distance teaching, organizational forms of students' work, recommendations for effective tutoring.*

---

### Aspects of Distance Learning in Ukraine

---

Distance Learning (DL) is being actively developed and becomes more and more popular all over the world. A wide range of citizens is interested in DL because of its accessibility. However, having some specific problems such as, for example, physical separation of students and teachers from each other, separation and mainly asynchrony of learning time, learning group's heterogeneity in knowledge and skills level, experience, geographical position and age, DL requires a serious research work.

Many scientists, especially in Ukraine, are very interested in information and communication technologies and their role in DL, they develop technical side of the question, different aspects of virtual learning environments usage, forgetting about tutor, who is a coordinator, leader, administrator of learning process, they forget about methodological side without which even an ideally equipped distance course (DC) will represent not more than an ordinary software or an electronic book. That is the **tutor's professionalism, personal qualities, methodological skills**, and also the **participants' interaction in DC** can make a distance learning process interesting, fascinating, lively, useful and efficient. To develop tutor's skills, to provide effective distance learning process, it is necessary to create a methodology for Distance Learning providing a harmonic interaction for participants of learning. That is why the research is devoted to Organizational Forms of Students' Learning (OFSL). **Aim of the research** is to create an effective methodology for students' distance learning for a distance course "Web-Design Technologies, 1 module". **Tasks of the research** are to examine the DL process from the point of view of OFSL and learning groups' homogeneity and heterogeneity, to analyse the received experimental results and to give practical recommendations on how to use the methodology in tutors' practice.

---

### Tutoring Methods in Distance Course "Web-Design Technologies"

---

**The main idea of the methodology** is to unite different students' organizational forms in one distance course. It is provided to have a possibility for **students to select** that or this OFSL for communication, which seems to be more convenient for them depending on their level of preparation, on the free time that they have, that can easily be put into practice technically and organizationally exactly in DL.

**Frontal-team form of learning**, as a rule, attracts **good students** more, who have experience and definite skills connected with the topic of DC and have something to share with other students. Frontal-team work is advantageous because of the fact that there are no limits in quantity of students and, so, the necessary variety of opinions and ideas is provided. This fact frightens **bad students**. They almost never address big group for help. Often their questions are addressed to the tutor. Probably, such behavior is connected with fear to write to students among whom there are, of course, good ones. It is connected with their lack of self-confidence, of knowledge and skills, with their inferiority complex, that later leads to a risk of ineffective learning. **Bad and average students** prefer working in small **teams** with 5-8 students being formed in a course as homogeneous in their level of knowledge.

However these two forms of work (frontal-team and team work) are not mutually exclusive but supplement each other in DL. Frontal-team work for bad and average students is a qualitatively new level of participation



in discussions, which is necessary to strive for. So, combining all OFSL helps to provide **individualistic approach to students**.

The suggested methodology does not limit students in selecting OFSL. Any of them can choose this or that form of learning, to take part or not to take it in this or that activity, to be in a role of supervisor or active participant that provides **democracy for distance learning**.

Besides, combining different OFSL is used to diversify the distance learning process, being mainly deprived of face-to-face contact and emotions. This helps students to see the process of learning as a many-sided and interesting activity, not as a boring pastime, that additionally **motivates distance students**.

---

### Practical Research Results

---

The correctness of this hypothesis has been checked on the created by the author a Distance Course "Web-Design Technologies, 1 module". From 2003 till 2005 on the basis of Research Laboratory of Distance Education in National Technical University "KhPI" (Kharkiv, Ukraine, <http://dl.kpi.kharkov.ua/techn/rle>) seven students' learning groups have been conducted in three different methodologies. In all three methodologies individual (making individual practical tasks) and frontal (supporting the learning process with announcements, reminders etc.) work were present. The first methodology (I) was different because teamwork was used there (making joint creative team-projects). The second methodology (II) included frontal-team students' work (constant asynchronous communication between students in big groups by mailing list). The third methodology (III) combined all the listed above organizational forms of work (individual, team, frontal and frontal-team work).

In total, 258 students from different cities of Ukraine, Russia, Bulgaria, Moldova and Estonia took part in the experiment. The working language was Russian, spoken by all the participants.

**Results analysis of experimental research** being hold by one of the authors [Molodykh, G., 2004] has shown that students who studied with combination of all OFSL (III) performed course tasks better ( $T_{\text{observed}} > T_{\text{critical}}$  (0,211 > 0,192)), and also were more active in communication by E-mail (834 letters), than students who studied either in teams (703 letters) or with frontal-team (392 letters) support [Bykov, V., Kukhareno, V., Molodykh, G., 2004]. The received experimental data allow talking about correctness of the hypothesis and, so, the methodology of OFSL combination is more effective than widespread and used in learning other methodologies (I and II). Besides, the learning results were later analyzed with a help of criterion  $\chi^2$  [Gnedenko B.V., Khinchin A.Y., 1971] that has also confirmed the correctness of the hypothesis.

---

### Recommendations for Tutors

---

So, there are some practical recommendations for tutors to help them in using the methodology of combining OFSL in distance learning.

**Preparation for Distance Learning.** Explain your students the essence of the individual work, teamwork, frontal and frontal-team work in one of the first tutor's letters and on the course web site. Many students, not having experience in distance learning, do not know what exactly is proposed to choose. Explanations can also include examples of students' functions after selecting this or that OFSL, how much free time is necessary for learning depending on OFSL.

Carry a preliminary questioning as soon as possible to find out more about your distance students: defining interests, wishes, technical and organizational potentialities of participants in a forthcoming learning process and finding out some personal data to build up approximate students' portraits.

Give students advice and written instructions about effective individual, team, frontal and frontal-team work, note the effective ways of interaction with other students.

**Individual work of distance students.** Give your students opportunity to study new learning materials individually on the course web site or with a help of additional learning materials in Internet, CD-disks or other storage devices. Do not forget that individual work is the main one, though, of course, not the only distance students' form of work.

Mastering the "core" of any new material, receiving reproductive and constructive skills should take place individually. Success of mastering a material and becoming proficient in elementary skills is an individual process,

which depends on many factors: amount of student's free time, individual student's characteristics (speed of reading, quickness of understanding, skills of analysis and synthesis etc.). Considering his/her own intellectual, organizational and technical opportunities, any student has a possibility to master the suggested material effectively. This is the peculiarity that advantageously differ distance learning from traditional face-to-face learning.

Announce about additional technical tools to perform practical tasks (open resources on the course web-site, sending tasks by e-mail etc.), about terms of performing tasks, point what and how should be sent to the tutor to be checked. Such announcements are easy to be sent by e-mail (by mailing lists), that is the main everyday tool of many students. Also place the information on the course web site.

Announce about evaluation and assessment of students' work on every type of tasks. Place the description of these criteria in the learning materials on the course web site, and also send them by e-mail.

Be operative in answering the letters of your students being written personally. Our experience shows that a delay with answer makes students worry.

At the end of the distance course let students work more at their projects individually. But be strict in explaining the deadlines because constant granting of delays provokes risks of students' unsuccessful finish.

**Frontal form of students' work.** This form is widely used in a face-to-face learning and is being transformed very much in distance learning. Some functions of teachers (for example, giving lectures), that take a lot of time in face-to-face learning, are being performed by electronic learning materials in DL. That is why frontal form of learning in DL is: giving announcements about forthcoming events, about learning results, reminders about deadlines etc.

As far as DL is almost deprived of synchrony, frontal form of work allows tutor to keep students informed of different events. Technically frontal work can easily be organized with the help of mailing list.

Try to give students regular reminders about the beginning of learning weeks listing practical tasks to be done and about deadlines to perform these tasks, about forthcoming discussions (date, time, terms etc.), about some changes on the web site (adding new learning materials, rating tables), information about final results of learning in a distance course etc.

**Team form of students' work.** Prepare a preliminary questionnaire for students to find out about their level of knowledge and skills, connected with course subjects and about personal data (sex, age, city, country, education etc.).

Depending on the results of questioning form small teams and suggest them to your students. Try to form teams homogeneous in preparation level and heterogeneous in sex, age and geographical position. This, on the one hand, provides comfort in communication while performing team projects and, on the other hand, makes the process of learning diverse and attractive with heterogeneity of a team.

Our practice shows that teams with 5 or less members will have organizational problems with communication as a result of rare contact between participants. And teams with more than 8 students will hardly find common points of view. That is why try to form small groups with 5-8 students.

Watch the process of work in teams. Intrude their work if it is necessary: noticing mistakes and inaccuracy while performing tasks, noticing and helping to solve conflicts.

Summarize teamwork results: make a competition, probably, with voting between participants of learning. Give comments on team projects, noticing successful decisions and recommending possible better approaches.

**Frontal-team form of students' work.** Initiate discussions on urgent subjects by mailing list. Prepare discussions beforehand; include participation of experts if possible. This will help to motivate students additionally and make the learning process more effective.

Pay students' attention to a possibility of initiating their own discussions in the course. Also explain students that they can have consultations not only with their tutor, but also with their colleagues in a big group. This will reduce additional tutor's working load and provide necessary experience and ideas exchange among students.

**Selecting organizational form of students' work.**

---

Activity of distance students is closely connected not only with their level of preliminary preparation but also with the amount of their free time for learning. Majority of distance students together with distance courses also study of work, that is why the occurring circumstances can be much unexpected, and tutor should be ready for them. That is why tutor should give students the opportunity to change the form of learning, the methodological combination during the learning process depending on the situation.

On the one hand, teamwork takes much more time, than individual; however, on the other hand, it develops valuable skills of teamwork and communication, which are necessary for any person in a modern world. Suggest those students who cannot operatively participate in teamwork take part in it passively, for example, to be still subscribed to the mailing list of the team and to be informed of any its events, but to perform the task, being given for this team, individually in any convenient time.

As it has already been said before, a student can choose the degree of activity in whole group discussions. However, try to add some students' questions, being asked individually but actual for the whole learning group, to the frontal-team communication (but only after their agreement!). Remember that a student's change of group from a small team to a big group is a qualitatively new level of his/her self-affirmation, knowledge and skills confirmation. That is why encourage such aspirations, compliment students for urgent questions and suggest all the learning group try to find answers to them.

---

### **Tutor's Diary**

---

A well-planned tutor's work is the first step to an effective distance learning process. Planning tutor's activities for every week allows him/her to carry out the duties logically, to help building the methodologically well-grounded learning process. Existence of strict plan allows tutor to perform the duties in time, which is very important in distance learning when separation in learning time is rather big, and to be more confident, making sure that everything being planned has been fulfilled and, so, to pay more attention to the richness of content in the learning process, not only to the organizational questions.

To solve this problem authors developed Tutors' Diary, which include a table from 7 columns, i.e. learning days in one week, and 4 rows, i.e. 4 organizational forms of student's learning that exist in DL. Such division can help to consider the distance learning process organization systematically.

The Diary has a banner where some data of current week are described (title and terms of distance course, tutor's name, week number and its length), key words (glossary), goals and tasks and also the basic knowledge necessary to master the material of the current week.

In Table 1 is shown an example of Tutor's Diary being filled in for the first week of the Distance Course "Web-Design Technologies".

Filling in the Diary tutor describes all the tasks that should be done during the week. So, one table is being made up per one week of learning. The Diary is especially useful during the first week of learning when many organizational and technical problems are being solved. The Tutor's Diary should always be "at hand" in a printed (or hand written) version that is very convenient and allows making changes quickly during the distance learning process. Besides, tutor can use the Diary, which has been approved in some distance course, repeatedly while tutoring in the same distance course again. This will save a large amount of time for preparation. In future the Diary can be included into a virtual learning environment as an additional working tool for tutors.

---

### **Conclusions**

---

The developed methodology of combining all organizational forms of students' learning allows providing individualized approach to students, to combine individual and collaborative type of learning, homogeneous and heterogeneous learning groups and to diversify considerably the distance learning process. Recommendations in using this methodology can help to approve it in other distance courses, because its positive moments are based on General Pedagogy appropriateness that were effective in the Distance Course "Web-Design Technologies".

Table. Tutor's Diary

Distance Course: <u>Web-Design Technologies</u> <u>Terms of learning: 10.11.03-21.12.03</u>		TUTOR'S DIARY Anna Molodykh Week # <u>1</u> Date: <u>10.11.03-16.11.03</u>		Key words: <u>Browser, hypertext, home page, list markers, marked list, unmarked list, tag parameters, definition list, tag, text editor, HTML</u> Goal, tasks: <u>Creating separate web-pages using hypertext markup language (HTML)</u> Basic knowledge and skills: <u>skills to work in Internet, with e-mail, in a text editor, Notepad</u>			
1. MON 10.11.03		2. TUE 11.11.03	3. WED 12.11.03	4. THU 13.11.03	5. FRI 14.11.03	6. SAT 15.11.03	7. SUN 16.11.03
Individual form of students' work	1. Introducing instructions on individual, team, frontal and frontal-team work. 2. Preliminary questioning.					<b>Deadline for performing tasks – 18.00.</b> 1. Checking, assessment of works, giving recommendations	
Students' teamwork		1. Forming teams. 2. Letter-proposal about forming teams.				1. Correcting formed teams.	
Frontal-team form of students' work	asynch.	1. Discussion «Please, introduce yourself!»	1. Discussion «Why to study HTML?»			1. Summarizing discussions.	
	synch.			1. Introductory Chat 1, 17.00.	1. Summarizing Chat 1 results.		
Frontal form of work	1. Invitation to study theoretical materials and performing tasks per Week 1	1. Invitation to Introductory Chat 1.	1. Invitation to go through the Test 1	1. Reminding of Deadline to perform tasks.			
Other	1. Opening access to learning materials 2. Place the information about students on web-site.		1. Open access to Test 1. 2. Subscription addresses to teams mailing lists			1. Update rating table on the web site	
Conclusions and notes: _____							

## Bibliography

- [Bykov, V., Kukhareno, V., Molodykh G., 2004] Bykov, V., Kukhareno, V., Molodykh G. Influence of Learning Process Organization on Students' Achievements in Distance Learning // The Fourth International Conference "INTERNET-EDUCATION-SCIENCE-2004", Baku-Vinnytsia-Veliko Tynovo, September 28 - October 16, 2004, V.1. - P.178-181.  
<http://anna-molodykh.narod.ru/pub-12.htm>

[Molodykh, 2004] Molodykh G. Individualna, grupova ta frontalna formy roboty studentiv v distantsijnomu navchanni (Individual, Team and Frontal Form of Students' Work in Distance Learning) - Information Technologies in Education, Science and Techniques / The 4<sup>th</sup> Ukrainian Conference of Young Researchers: April 28-30, 2004, Cherkasy, Ukraine – Pp.31-34. <http://anna-molodykh.narod.ru/pub-11.htm>

[Gnedenko B.V., Khinchin, A.Y., 1971] Elementarnoe vvedenie v teoriyu veroyatnostei (Elementary introduction to the Theory of Probability) – Moskow, 1971.

---

### Authors' Information

---

**Valeriy Bykov** – Doctor, Director of Educational Environments Institute in Educational Sciences Academy of Ukraine. Address: Berlinsky st., 9, Kyiv, Ukraine; e-mail [bykov@edu-ua.net](mailto:bykov@edu-ua.net)

**Yuriy Zhook** – Ph.D. in Pedagogy, Assistant Director of Educational Environments Institute in Educational Sciences Academy of Ukraine. Address: Berlinsky st., 9, Kyiv, Ukraine; e-mail: [zhook@edu-ua.net](mailto:zhook@edu-ua.net)

**Ganna Molodykh** – Post-graduate student of Educational Environments Institute in Educational Sciences Academy of Ukraine; a senior lecturer of Cross-Cultural Communication and Modern Languages Department in National Technical University “Kharkiv Polytechnic Institute”. Address: Barabashova st. - 38-A, apt.47, Kharkiv, Ukraine, 61168; e-mail: [molodykh@kpi.kharkov.ua](mailto:molodykh@kpi.kharkov.ua) [molodykh@ukr.net](mailto:molodykh@ukr.net) web: <http://anna-molodykh.narod.ru>

## УЧЕБНАЯ МОДЕЛЬ КОМПЬЮТЕРА КАК БАЗА ДЛЯ ИЗУЧЕНИЯ ИНФОРМАТИКИ

**Евгений А. Еремин**

***Аннотация:** Предлагается в качестве общей основы изучения различных дисциплин курса информатики, связанных с вопросами software и hardware, использовать единую учебную модель компьютера. Дается обоснование данного подхода, перечислены разделы, где он является наиболее актуальным. Описаны собственные разработки автора по проблеме (включая программную поддержку) – учебная модель компьютера "E97" и компилятор с языка Паскаль для нее. Подчеркивается, что обсуждаемые идеи пригодны и для других учебных моделей.*

***Ключевые слова:** модель компьютера, обучение, информатика, компилятор.*

---

### Введение

---

Одной из существенных принципиальных трудностей в изучении курса информатики является комплексный характер содержания этой молодой и быстро развивающейся науки. В частности, сам компьютер, будучи по своей сути неразрывным единством различных технологий, в образовании несколько искусственно разделяется на software и hardware (см., например, обзор компьютерных учебных предметов в [1]). Дистанция между этими двумя дисциплинами имеет тенденцию к увеличению: многочисленные программные слои (ROM BIOS, операционная система, языки высокого уровня и визуальные системы, прикладное ПО) все больше и больше отделяют работающих на компьютере от его аппаратной части. Калейдоскопическая смена моделей hardware и версий software еще более осложняет подбор материала для обучения.

В результате при традиционном содержании курсов мы имеем пользовательский интерфейс и основы языка высокого уровня с одной стороны, и вычислительную технику с ее абсолютно "невидимой" на практике двоичной системой с другой. Разумеется, хороший лектор обязательно демонстрирует и всячески подчеркивает их связь, но далеко не все студенты осознают это реально существующее

неразрывное единство. К сожалению, с каждым годом указанное разделение обостряется, что осложняет формирование правильного мировоззрения наших обучаемых.

Один из возможных путей логического объединения содержания курсов, связанных с software и hardware, состоит в их изучении на некоторой общей основе. Поскольку реально существующие компьютеры имеют сложное, сильно различающееся в деталях и быстро меняющееся устройство, при построении наиболее глубоких курсов часто используется прием, который заключается в замене компьютера учебной моделью [2-5]. Такие модели наглядны и просты для понимания с одной стороны, и сохраняют все наиболее важные и неизменные свойства реальных машин с другой. По-видимому, модель MIX известного математика Доналда Кнута, предложенную как некоторую абстрактную базу для изучения фундаментальных закономерностей программирования и вычислений на ЭВМ [2], можно отнести к наиболее обстоятельным. Заметим, что недавно автор указанного классического труда по компьютерным вычислениям обновил свою модель [6,7], сделав ее гораздо более современной.

Целью данной работы является стремление показать, что при изучении курса информатики учебная модель компьютера может использоваться гораздо шире и фактически играть роль связующего звена при изучении аппаратной и программной частей. Хотя автор первоначально использовал собственную учебную модель "E97" [8,9], все излагаемые идеи вполне могут быть реализованы и на других моделях.

---

### **Возможные применения учебных моделей при изучении информатики**

---

Наиболее очевидным разделом, где можно вести изложение материала на базе учебной модели, является устройство ЭВМ. Поскольку хорошая модель обязательно отражает все наиболее существенные черты оригинала, данное положение в примерах и доказательствах не нуждается.

Другой подходящей темой являются системы счисления, где помимо знакомства с общей теорией можно на практике реализовать алгоритмы перевода, например, из десятичной системы в двоичную и обратно. Формы использования модели могут зависеть от педагогических целей: разработка программы перевода, разбор готовой программы, знакомство с ее работой и тестирование для некоторых интересных случаев (знаковые и беззнаковые числа, отрицательные значения, большие числа и переполнение разрядной сетки и т.п.). В результате вместо скучного абстрактного перевода чисел на листке бумаги появляется возможность опробовать эти процессы на практике.

Развивая предыдущую идею, можно предложить на базе учебной модели изучать кодирование и нечисловых видов информации – текстов и графики. При этом попутно обязательно будут рассмотрены вопросы байтовой структуры памяти, объединения байтов в многобайтовое значение (big или little endian), кодировки ASCII, Unicode и другие, а также множество аналогичных фундаментальных проблем.

Интересным применением служит также подробное изучение на модели выполнения логических операций И, ИЛИ, исключающее ИЛИ, НЕ и др. Это тем более важно, поскольку помимо высокоуровневых логических операций над переменными типа Boolean, широко применяемых в различных условиях и запросах, существуют еще и поразрядные одноименные операции над двоичными кодами.

Наконец, та же самая учебная модель компьютера может служить базой и для изучения программного обеспечения. В первую очередь, это фундаментальные идеи работы компиляторов (ниже это будет продемонстрировано), а также программ обработки текстов, архиваторов и других видов прикладного ПО. Наглядный разбор на модели даже нескольких простейших случаев (преобразование регистра у символов, переход от числа к его текстовому представлению и обратно, сжатие повторяющихся последовательностей кодов, преимущества неравномерного кодирования) часто улучшает понимание принципов обработки информации на ЭВМ существенно быстрее, чем несколько абстрактных деталей лекций или решение многочисленных пользовательских задач с помощью прикладного ПО.

При наличии достаточно развитой модели можно даже попытаться продемонстрировать внутреннюю логику параллельных вычислений или многозадачного режима.

Таким образом, главным новшеством в предлагаемом подходе является не идея применения учебного компьютера как основы изучения устройства компьютера, но возможность демонстрации на базе единой модели ЭВМ других тем курса информатики. В частности, подобный подход для изучения принципов функционирования программного обеспечения в настоящее время практически не используется.

---

### **"E97" как пример базовой модели**

---

Как уже подчеркивалось выше, сам по себе выбор в пользу той или иной учебной модели не является принципиальным. Тем не менее, применяемая модель должна как можно лучше соответствовать содержанию курса информатики, быть простой и наглядной, и, в то же самое время, отражать большинство характерных черт современных компьютеров. Следовательно, в качестве единой базы для комплексного курса подходит не каждая модель.

Не претендуя на единственность, в 1997 году автором была разработана учебная модель компьютера "E97" [8]. Позднее ее устройство было изложено в книгах, вышедшим большим тиражом [4,9]. Модель используется в преподавании различных тем в целом ряде учебных заведений России.

При создании в "E97" были заложены следующие особенности, отличающие ее от других моделей:

- соответствие реальным принципам устройства персональных компьютеров (байтовая организация памяти, обмен через порты ввода/вывода и т.п.);
- максимально простая по логике, но развитая система команд;
- возможность работы с нечисловыми данными, причем занимающими разный объем в ОЗУ;
- современная адресация к ОЗУ, в частности, косвенная адресация через регистры процессора;
- широкое применение библиотеки готовых подпрограмм, которые дополнительно являются тщательно разработанными образцами программирования;
- существование нескольких возможностей уровней освоения, начиная с ознакомительного и кончая полноценным программированием на языке процессора.

"E97" – это учебная модель компьютера, по принципам архитектуры во многом похожая на известное своей строгой логичностью и наглядностью семейство компьютеров PDP-11 [10]. Модель содержит процессор, память двух видов – ОЗУ и ПЗУ (постоянное запоминающее устройство), а также реализует имитацию аппаратной работы с клавиатурой и дисплеем. Она способна обрабатывать как числовую (двухбайтовую), так и текстовую (однобайтовую) информацию.

Важной особенностью описываемой модели является тщательно разработанное ПЗУ. Его наличие существенно облегчает общение обучаемого с внешними устройствами, фактически сводя его к вызову стандартных подпрограмм. Указанный прием в реальной вычислительной технике также имеет место: в компьютерах IBM PC, например, подобное ПЗУ носит название ROM BIOS [11]. С образовательной точки зрения немаловажно, что содержимое ПЗУ, представляющее собой в реализации подробно прокомментированный текстовый файл, само способно служить материалом для изучения.

Как и всякая многоуровневая модель, "E97" допускает широкую индивидуализацию сложности заданий, что позволяет ее применять на занятиях в учебных заведениях различного уровня.

---

### **Демонстрационный компилятор на базе учебной модели**

---

Учебная модель может служить основой и для знакомства с принципами работы программного обеспечения. Важность такого подхода во многом следует из способа освоения вычислительной техники на современном этапе. Дело в том, что компьютерные специалисты с большим стажем совершенствовали свои знания по мере развития вычислительных машин. Это дало им возможность естественным образом перейти от процессорных кодов к современным языкам высокого уровня, освоив тем самым полную палитру способов программирования. В настоящее время подавляющее число людей начинают обучение в лучшем случае с языков высокого уровня (а то и вовсе игнорируют программирование, из-за чего логика работы машины для них остается тайной). В результате из-за отсутствия хотя бы минимального знакомства с работой процессора, многие концепции, такие как способы передачи параметров или экономичные методы построения структур данных, оказываются не до конца понятыми.

Для компенсации описанного выше недостатка автор предлагает использовать разработанное им специализированное учебное программное обеспечение, демонстрирующее наиболее важные принципы автоматического составления программ. Очевидно, что профессиональные компиляторы для подобных целей подходят плохо, поскольку они абсолютно закрыты и генерируют "тяжелый" для разбора код.

Учебный демонстрационный компилятор "КомПас" оперирует с некоторым ограниченным подмножеством языка Паскаль. Он включает все алгоритмические структуры: условный оператор IF и традиционные циклы WHILE, REPEAT и FOR. Стандартные процедуры ввода и вывода READ/WRITE реализуются через вызов подпрограмм стандартной библиотеки, которые, в свою очередь, после необходимой подготовки переадресуют обращение к ПЗУ. Компилятор поддерживает стандартные типы данных и массивы из них.

Перечисленные возможности позволяют продемонстрировать студентам следующие существенные черты языков высокого уровня:

- переменные, константы, типизированные константы и разница между ними;
  - организация хранения данных разного типа в ОЗУ, включая массивы, и методы доступа к ним;
  - преобразование значений разных типов данных (например, CHAR в INTEGER и т.п.);
  - способы реализации базовых алгоритмических структур;
  - механизмы использования процедур
- и другие фундаментальные принципы.

В качестве иллюстрации рассмотрим пример простейшей программы на Паскале:

```
PROGRAM sample;
CONST x = 2;
VAR y: INTEGER;
BEGIN y := x + 10;
      WRITELN(y)
END.
```

В результат трансляции "КомПас" сгенерирует короткий и наглядный код, представленный на экране в виде таблицы с детальными комментариями. Приведем ее для случая процессора Intel, поскольку это не потребует дополнительных пояснений (хотя код для учебной модели получается более простой, тем не менее, для его прочтения в статье потребовалось бы более подробное описание).

Адрес	Код	Ассемблер	Действия	Комментарий
100	E97D01	jmp 0280		переход на начало
103	...			библиотека стандартных программ
280	B80200	mov ax,0002	2 ==> ax	константа x
283	B90A00	mov cx,000A	10 ==> cx	константа 10
286	01C8	add ax,cx	ax + cx ==> ax	x + 10
288	A3FE04	mov [04FE],ax	ax ==> [4FE]	сохранить результат в y
28B	A1FE04	mov ax,[04FE]	[4FE] ==> ax	загрузить значение y
28E	E8C6FE	call 0157	печать integer	WRITE y (обращение к библиотеке)
291	E8FDFE	call 0191	на след. строку	LN (обращение к библиотеке)
294	CD20	INT 20	выход в систему	END

Анализ полученной программы особых трудностей не представляет и вполне доступен даже начинающим. Учебный компилятор свободно распространяется через Интернет и может быть получен с Web-страницы [12]. Там же можно найти ссылки на подробную on-line документацию.



---

## Заключение

---

Таким образом, идея использования единой учебной модели компьютера в качестве основы для изучения различных разделов курса информатики оказывается применимой к целому ряду тем. Указанный подход, как свидетельствует педагогический опыт, позволяет существенно повысить качество знаний и формирует более адекватное представление об обработке информации средствами современной вычислительной техники.

---

## Ссылки

---

- [1] Computing Curricula 2004 (draft). URL: <http://www.acm.org/education/curricula.html>
- [2] Кнут Д.Э. Искусство программирования. – Москва: Издательский дом "Вильямс", 2000. (Donald E. Knuth. The Art of Computer Programming. Reading, Massachusetts: Addison-Wesley, 1997)
- [3] Брукшир Дж. Гленн. Введение в компьютерные науки. Общий обзор. – Москва: Издательский дом "Вильямс", 2001. (J. Glenn Brookshear. Computer Science: an overview. Reading, Massachusetts: Addison-Wesley, 2000)
- [4] Могилев А.В., Пак Н.И., Хеннер Е.К. Информатика. – Москва: Академия, 1999.
- [5] Методика преподавания информатики. Учеб. пособие для студ. пед. вузов / М.П. Лапчик, И.Г. Семакин, Е.К. Хеннер; Под общей ред. М.П. Лапчика. – Москва: Академия, 2001.
- [6] Knuth D.E. MMIXware: a RISC Computer for the Third Millennium. Heidelberg, Springer-Verlag, 1999.
- [7] MMIX Homepage. URL: <http://www-cs-faculty.stanford.edu/~knuth/mmix.html>
- [8] Еремин Е.А. Как работает современный компьютер. – Пермь: издательство ПРИПИТ, 1997.
- [9] Еремин Е.А. Популярная лекция об устройстве компьютера. – Санкт-Петербург: BHV-Петербург, 2003.
- [10] Лин В. PDP-11 и VAX-11. Архитектура ЭВМ и программирование на языке ассемблера. – Москва: Радио и связь, 1989. (Wen C. Lin. Computer Organization and Assembly Language Programming for the PDP-11 and Vax-11. NY: Harper & Row Publishers)
- [11] Нортон П. Персональный компьютер фирмы IBM и операционная система MS-DOS. – М.: Радио и связь, 1992. (The Peter Norton Programmer's Guide to the IBM PC. Microsoft Press)
- [12] ComPas Homepage. URL: <http://www.pspu.ru/personal/eremin/eng/myzdsoft/compas.html>

---

## Информация об авторе

---

**Евгений Александрович Еремин** – Пермский государственный педагогический университет. Россия, 614990, Пермь, Сибирская, 24. e-mail: [eremin@pspu.ac.ru](mailto:eremin@pspu.ac.ru)

## HISTORICAL INFORMATICS IN PRACTICAL APPLICATION: VIRTUAL MUSEUM OF THE KHAZAR STATE HISTORY

**Boris Elkin, Alexander Kuzemin, Alexander Koshchy, Alexander Elkin**

### *Extended Abstract:*

The process of informatization having touched just about all areas of knowledge involved the historical science in the same manner between eighties and nineties of the XX<sup>th</sup> century. Computers have become not only more accessible to historians, technical characteristics of personal computers and their software are improved steadily, this transforms a computer into more and more attractive and effective instrument of historical investigation. It creates real possibilities for application of the newest information technologies to the work of a historian.

Rapid growth of the computer industry resulted in creation of the International History and Computing Association in 1986, it coordinates activity of the historians from different countries who apply computer methods and technologies to their research practice and educational process. Since 1992, the scientific centers from NIS have joined the Association.

The processes typical for the nineties consisting in the rise of the branch informatics in a number of the fields of science were the reason for creation of the Association. One of such new branches of science emerging on the border of the information science and social sciences and humanities is *the historical informatics*.

***Historical informatics*** is a branch of science, which investigates regularities of the historical science and education informatization process; the basis for the historical informatics is the totality of theoretical and applied knowledge necessary for creation and use when studying in practice all kinds of historical sources electronic versions.

The recent concept of information including the social information and theoretical source studies is *the theoretical basis of historical informatics* and information (computer) technologies are *the basis of the applied historical informatics*.

*The sphere of interests of the historical informatics* incorporates development of the general approaches to application of information technologies to historical investigations, in particular, specialized software; creation of historical data bases and banks (knowledge); application of information technologies to data presentation and analysis of the structured, text, image and other sources; computer simulation of historical processes; use of information networks (Internet and others); development and use of multimedia means and other modern directions in the historical science informatization; application of information technologies to the historical education.

The historical informatics is a young science in Ukraine. In addition, it is quite natural that in its development and formation it meets with some difficulties. There is a good reason to single out among the main problems the following ones:

- a). Teaching of the disciplines associated with application of new methods to the historical investigations is performed in one form or another only in some cities of Ukraine. There are several reasons for this and, first, it is *the lack of computers and specialists*. Computers are still not easily accessible for the humanitarian faculties.
- b). The lack of *the specialization in the given direction* in the majority of universities both on the level of the diploma theses and dissertations limits the possibility to train experts.
- c). There is a constant *deficiency of valuable periodical editions* oriented to the historical informatics.

Solution of the above-mentioned problems would make it possible to rise the level of investigations in the historical informatics in our country.

It should be noted that the *Western branch of the International Solomon University* having taken an interest in this promising scientific direction has much potential to solve the above-mentioned problems in many instances at the expense of the resources of its professional, methodological, material basis and links with higher educational establishments in NIS and abroad.

Among the first experimental projects it is scheduled to create the specialized software aimed to apply computers to perform work carried out in the *International Center of Khazar Studies* which is a part of the University in the right of research subdivision from 1999.

The main *concern of the Center* consists in organizing and performing fundamental and applied scientific researches in the history of the Khazar Kaganat<sup>1</sup>. The base for carrying out such researches is a great burial ground of the Khazar period located in Kharkov region; this burial ground contains about 20 thousand unexplored burials.

A great number of the accumulated findings and unique exhibits, the necessity for revealing various regularities and interrelations of the material under investigation, the necessity to use computers for improvement and perfection of research processes carried out in the Center justify the idea of creation of the specialized software combining various achievements in the area of modern information technologies. The availability of the present-day computer faculty and the best experts in the area of computer technologies in the city make it possible to

---

<sup>1</sup> The Khazar Kaganat is one of the greatest and the most developed states at VII-X A. D. located on the territory from the Volga to the Dnieper and from Russian-Ukrainian borders, Kharkov region among them, till the Crimea including.

---

perform the work on creation of the virtual educational research museum of the Khazar Kaganat history based on the Center of Khazar Studies as the first practical experience on the way of development of the historical informatics on the scientific grounds of Ukraine.

---

### Authors' Information

---

**Elkin B.S.** – Prof. Director , East Ukrainian branch, International Solomon University

**Kuzemin A.Ya.** – Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, (Ukraine), [kuzy@kture.kharkov.ua](mailto:kuzy@kture.kharkov.ua)

**Koshchy A.F.** - The senior lecturer, East Ukrainian branch, International Solomon University

**Elkin A.B.** - Bachelor, East Ukrainian branch, International Solomon University

## DISTRIBUTED INFORMATION MEASUREMENT SYSTEM FOR SUPPORT OF RESEARCH AND EDUCATION IN OPTICAL SPECTROSCOPY

**Sergey Kiprushkin, Nikolay Korolev, Sergey Kurskov, Natalia Nosovich**

**Abstract:** *The present paper is devoted to the distributed information measurement system for support of research and educational process with remote access to informational and technical resources within the Intranet/Internet networks. This system is characterized by the network integration of computer-based research equipment for the natural sciences. It provides multiple access to such resources in the networks functioning on the basis of Protocol Stack TCP/IP. The access to physical equipment is realized through the standard interface servers (CAMAC, GPIB), the server providing access to Intel MCS-196 microcontrollers, and the communication server, which integrates the equipment servers into uniform information system. The system is used for making research task solutions in optical spectroscopy, as well as for supporting educational process at the Department of Physics and Engineering of Petrozavodsk State University.*

**Keywords:** *distributed information measurement system, equipment server, CAMAC server, GPIB server, client-server technology, distance education.*

---

### Introduction

---

Widespread information packets (for example, "National Instruments" packets – LabWindows/CVI, LabView, BridgeView, and also the software systems of visualizing measurement information (SCADA-systems), provide more or less remote access to physical equipment. However, in these software tools the equipment is connected to the computer where the instrumentation packet is installed. It makes difficult the use of different interfaces connected with separate sub-systems and attached to separate computers. Besides, such software packets, though they have friendly interfaces and visual programming tools, do not assure the flexibility and expandability of the system structure. Therefore, they are not always effective for the systems of data collecting and experiment control.

As an example of a system with similar functions, one can take the data collecting system MIDAS (<http://midas.psi.ch/>), developed at Paul Scherrer Institute (Switzerland). This system is aimed at remote collecting of experiment data in nuclear physics and physics of elementary particles. However, the remote access within the MIDAS system is based on HTTP-protocol combined with the mechanism of distance procedure call. The advantages of this system are the lack of system monitoring and the inability to reserve any object for operation, which is necessary in a multi-user system.

The purpose of this work is to develop the distributed information measurement system for support of research and education in optical spectroscopy.

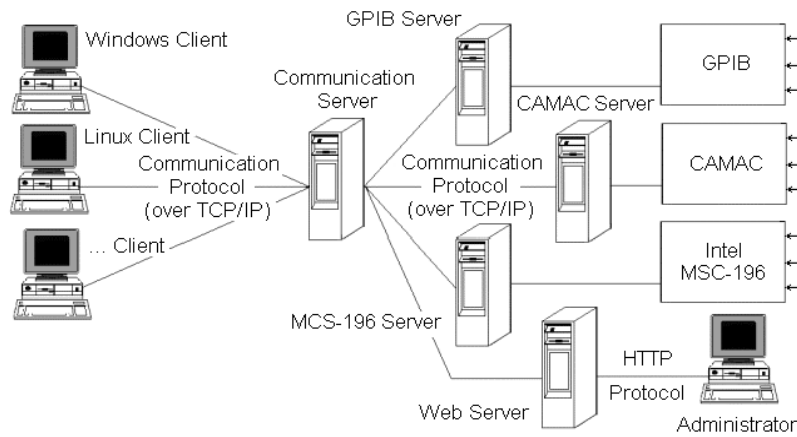
**The Distributed Information Measurement System**

The distributed information measurement system is built as a centralized system. See its scheme on figure 1.

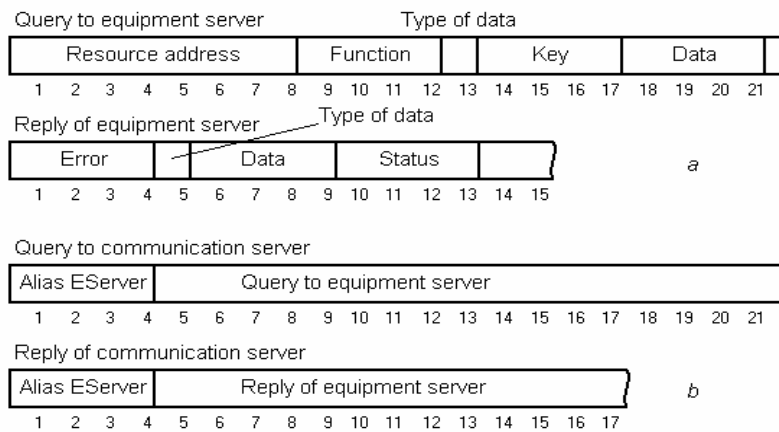
The system is comprised of the following parts: the communication server; the equipment servers (CAMAC server [Zhiganov et al, 2000], GPIB server [Gavrilov et al, 2003] and the server of MCS-196 microcontrollers); the client programs fulfilling the collection, accumulation and processing of information and experiment control; the universal protocol connecting the communication server with the equipment servers; and the extended protocol connecting the communication server with the client programs.

The program modules of the distributed system are realized in the Java programming language by means of TCP sockets technology. The methods of using the input-output ports for the access to the interface controllers are written in C programming language.

The data frame of the multi-equipment communication protocol between the communication server and the equipment servers and the data frame of the exchange communication protocol between the communication server and the clients are presented on figure 2.



**Figure 1.** The scheme of the distributed information measurement system



**Figure 2.** (a) The data frames for multi-equipment protocol between the communication server and equipment servers  
 (b) The data frames for exchange protocol between the communication server and the client programs

The request data frame for the equipment servers includes the following fields: the resource address – 8 bytes, the function – 4 bytes, the data type – 1 byte, the key – 4 bytes, the data – 4 bytes. If the "data type" field

is equal to zero, the "data" field contains the exact data. In case the "data type" field is equal to 1, the "data" field contains the length of the data in bytes message (the data directly follow the main frame). The "key" field is intended for system administration.

The reply frame from the equipment server contains the error number – 4 bytes, the data type – 1 byte, the data – 4 bytes, and the system status information – 4 bytes. The "system status information" field is used by the equipment server to transmit the contents of status registers to the user. The contents of these registers are devices-specific. The data frame for exchange protocol between the communication server and the client programs has an additional field (4 bytes), which contains the requested equipment server alias.

The equipment server has a generalized structure. Device interfaces differ from each other only in their methods libraries that realize the interaction between the equipment server and a certain device. The equipment server is a server for sequential request processing. It determines whether the requested function and the equipment address are admissible for this equipment, transmits the request and the address to the equipment and forwards the reply or the error code to the client program if any exceptional circumstances occur. So, this communication server, being a multipurpose one, handles equipment servers using the same protocol.

We will describe the structure of the equipment server taking the CAMAC server as an example. The CAMAC server includes the following classes:

- CamacS – the basic server class – fulfills "network listening" and connects the communication server; it also realizes the client service procedure – for example, checks the client's request input parameters and executes the device orders;
- CServerProtocol – this interface defines the operation codes, the error codes, and other constants of the communication protocols (common for the system);
- QueryToEServer – this class defines the "server request data frame" and the methods for working with it;
- ReplyFromEServer – this class defines the "reply data frame" and the methods for working with it;
- CamacLib – this class contains the methods library for working with CAMAC equipment. For reading and writing operations in the in/out ports, it refers to the external methods, which are written in C programming language. The basic class methods: CmZ – nonaddressable CAMAC operation zero, CmC – nonaddressable CAMAC operation clear, CmF – addressable control CAMAC operation, CmFW – addressable writing CAMAC operation, CmFR – addressable reading CAMAC operation, CmQ – L-request test operation. Also, the two external methods are included in this class: outport – byte writing – and inport – byte reading from the in/out ports.

The communication server functions are: providing multi-user mode support; resources distribution; the prevention of unauthorized access to the experimental equipment; system monitoring; and security control. The multi-user mode is implemented by means of parallel processes together with the synchronization of some functions. The system monitoring function includes keeping and providing (at the request of the administrator) information about the users working with the equipment at the moment. The system safety is provided by enciphering of traffic between the communication server, the equipment servers, and clients (see below). We consider it's necessary, i.e. there is a possibility that the data frames might be substituted by unauthorized users, which might lead to unauthorized access to the equipment of the system and to the information of clients who work with this system.

The communication server consists of the following general classes:

- StartCServer – server initialization. In this class, the survey of the all equipment servers is realized, as well as the connection with all the working equipment servers is established. After that the communication server proceeds to the mode of waiting for user connection;
- ServerThread – the class, which realizes the client service procedure. This class includes processing client commands, making requests to the equipment servers, refreshing information about the used modules (devices) connected to the corresponding device interface. The appeal to the in/out ports procedures is fulfilled from the critical section;

- CServerProtocol – the interface determining the operation and error codes, as well as other protocol constants;
- MainServInf – the class that keeps information about all the equipment servers in this distributed system. These data are stored as a set of records, which contain the following fields: the equipment server IP address, the equipment server port number for the communication server connection, the communication server alias, the socket (if the connection is established), and the equipment server status;
- MainClientInfo – the system monitoring class used for storing information about the system users working with the experimental equipment at the moment. The information contains the client IP address, the client identification number, and the client occupation resources (the equipment server alias and the resource address).

The ReplyFromEServer, QueryToEServer, ClientReply, ClientQuery classes define the "query frame" and the "reply frame" for the corresponding protocols of the exchange methods for the communication server and its client.

The communication server works in the following way. On starting, the communication server reads information about the available equipment servers from the configuration file (IP address, port number, alias). Then the communication server connects to all the equipment servers. If the connection with a certain server has not been established, this server is marked as unavailable at the moment. The server will repeat the connection attempt to the equipment server in case of any user's request to the device. After the initialization of the equipment servers, the communication server enters the client request-waiting mode. When a client connects to the server, it starts the parallel process for the client service and confers to the process a unique number CID (Client ID), which is not equal to zero. The server-client data exchange is realized in the "query-reply" mode by the expanded protocol. Both the query and the reply frames are the equipment server frames with the equipment server alias (4 byte). At the moment the following functions are available: CS\_GETRESOURCE – resource capture, CS\_RELEASERESOURCE – resource release, CS\_QUIT – work stop. To be able to work with equipment, a client has to reserve a certain resource by CS\_GETRESOURCE command noting the resource address and the equipment server alias. On the client demand, the communication server makes request to the equipment server with CS\_CHECKRESOURCE function for the resource address testing. If no errors occur, the communication server provides the resource to the client. The equipment access system uses hierarchical addressing system (for CAMAC equipment: the crate number, the station address, and the subaddress); the complete address is 8 bytes long. The client software sets the addresses into these 8 bytes, while the equipment server fulfills the inverse operation. After the resource capture procedure, the client can start working. The CS\_QUIT command stops the connection. All the used functions (the service functions) have numbers FFFFFFFFh and lower. The functions for the use of equipment (the device functions) have numbers with beginning from zero. The error codes range from zero to FFFFFFFFh. If any connection to the equipment server is broken off during the work, the communication server will attempt to reconnect after the client's next request.

---

## Data Security

---

Information security in the system is based on cryptographic JCE 1.2 extension from Java 2 Platform Standard Edition v1.4 packet and Cryptix 3.2. JCE 1.2 and Cryptix 3.2 provide the basis for developing encryption algorithms, key algorithms, and authentication algorithms. These software packets also provide many realizations of popular cryptographic algorithms.

All participants of the inter-net exchange have in their disposal the following classes:

- The generator of the key pair (the public key and the private key). In the system, the keys of RSA algorithm are used in RAW encoding, 1024 bigits in length. The keys are stored in files;
- The classes of the encryption and decryption of a random byte chain according to RSA algorithm. An encoded chain has a length of 128 bytes;
- The classes of the encryption and decryption of a random byte chain according to Rijndael algorithm. The encryption method is CFB (Cipher Feedback).

It means that the plain text is encrypted by portions up to 64 bytes long. At that the preceding portion of the cipher text is joined with the next portion of the plain text by "exclusionary OR" operation. In the system the encryption is realized by one-byte blocks;

- The class of digital signature (RSA algorithm and digest MD5 algorithm are used). The digital signature is designed for the authentication of the source of the message, the determination of the message's integrity, and the provision of the impossibility of the refusal from the fact of signing a certain message. The digital signature is calculated for a random byte chain. The signature is 128/129 bytes long;
- The class of the message digest calculation (MD5 algorithm is used). The digest has a length of 128 bigits. The message digest (the hash function) guarantees the data integrity;
- The class of MAC-code (message authentication code – MAC). HMAC-MD5 algorithm is used; the MAC-code is 128 bigits long. The message digest, which is calculated by means of encryption with a secret key. The MAC-code works as a digital signature with the system working in the protected mode.
- The class of the encryption and decryption of the files with the keys, based on the password set by the user. In this class, PBEWithMD5AndDES algorithm is used (JCE 1.2 specification). The encryption method is CBC (Cipher Block Chaining) – the encryption is realized by 8-bytes blocks, at that the block of the encrypted message, made as a result of encryption  $n$ -block of the plain text, and the next  $n+1$  block of the plain text are joined by "exclusionary OR" operation. Consequently, the next  $n+1$  block of the cipher text is made up.

Besides, the communication server has an additional class for generating the secret keys. The system uses 128-byte keys based on Rijndael symmetric algorithm. These keys play the roles of both session and temporary keys. The administrator of the communication server may vary key lifetime and thus limit the session time, or he may fix key lifetime for a specific period.

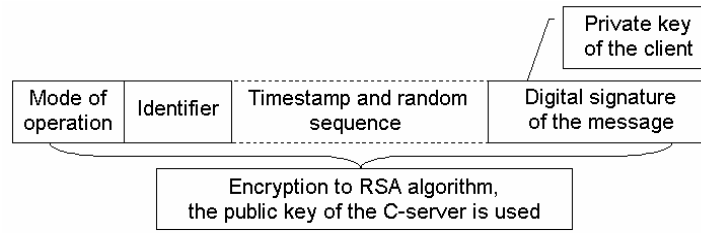
Each side of the inter-net exchange generates a pair of keys. All public keys are sent to the communication server, which, in its turn, provides its public key to the users and the equipment servers. The public keys are placed on the system administrator site for the revision of the received keys.

In the security system, there are three work modes that depend on how important the tasks for solution are. The first mode is public work in the system when the cryptographic security is completely switched off. This mode is for testing within the system and its checkout. The second mode is real work, which requires the integrity provision and the guarantee of the non-cancellation of data, and also the authentication of access rights, but no confidentiality. And the last mode is private work when all data are transmitted only encrypted with a digest or a digital signature. Here confidentiality is of greatest importance.

Now we will analyze the last two modes in detail.

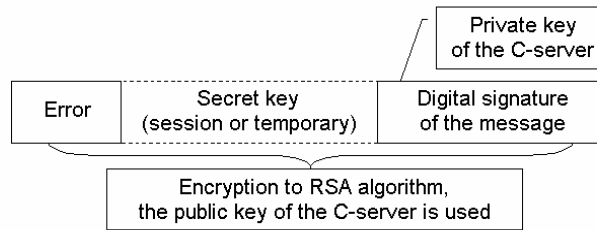
While initializing, the communication server carries out the decryption of the key store. The key store contains private and secret (temporary) keys. The administrator must provide correct password to start decryption. The communication server checks validity of temporary keys and changes them if their lifetime is over. Different secret keys are used for the users and for the equipment servers. The unique identification number given to the clients and the equipment servers allows defining who has which key. The information about the distribution of the rights of access to the equipment servers is stored in a special file. Each equipment server has a list of those clients who have access to it. The authentication of the rights of access is carried out by the communication server after the stage of identification and the secret keys checkout before making requests to the equipment server.

After the initialization, the communication server checks whether the equipment servers are ready for work. The equipment server, when connected to the communication server, replies with encrypted message which contains the identifier, the timestamp, and the digitally signed random sequence. The encryption is made by means of the public key of the communication server. Then the communication server performs the decryption and verification of the digital signature. If everything is all right, the timestamp is checked. If the specified timestamp corresponds to predefined interval, new session key or temporary key is sent to equipment server. The key is accompanied by the message about successful transition. In the other case, the communication server sends error message with correspondent timestamp and the random sequence. The reply of the communication server is encrypted by the public key of the equipment server and digitally signed by the communication server. The next attempt to establish connection will be made upon any user's request to the equipment server.



*Mode of operation – 1 byte;  
 Identifier – 4 bytes;  
 Timestamp and random sequence – 191 byte;  
 Digital signature – 128/129 bytes.*

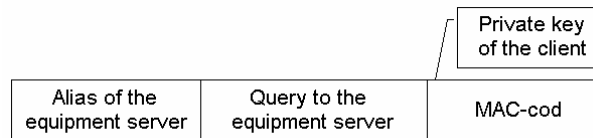
a)



*Error – 4 bytes;  
 Secret key – 191 bytes Rijndael algorithm. Instead of the key there may be timestamp and random sequence in the case of any error;  
 Digital signature – 128/129 bytes.*

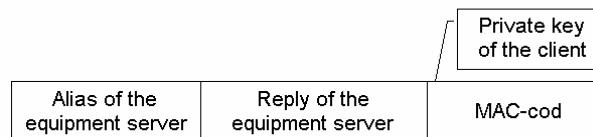
b)

**Figure 3.** The first message of the client for the communication server (C-server) (a) and the reply of the communication server (b)



*Alias of the equipment server – 4 bytes;  
 Query to the equipment server – 21 bytes;  
 MAC-code – 16 bytes.*

a)

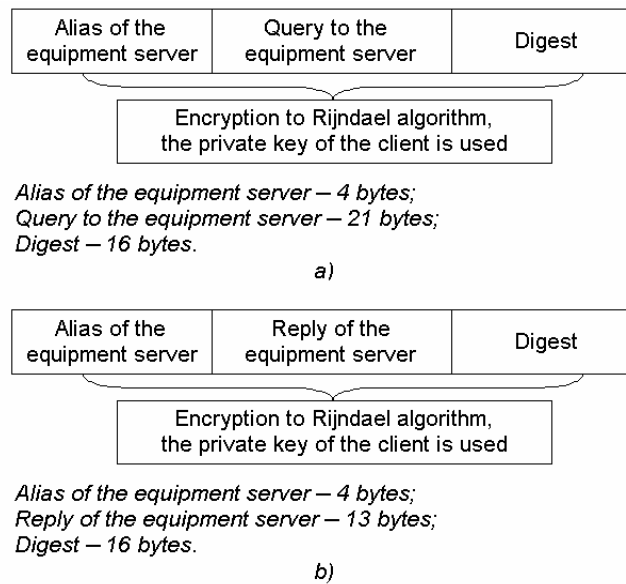


*Alias of the equipment server – 4 bytes;  
 Reply of the equipment server – 13 bytes;  
 MAC-code – 16 bytes.*

b)

**Figure 4.** The query of the client to the communication server in protected mode (a) and the reply of the communication server (b)





**Figure 5.** The query of the client to the communication server in the private mode (a) and the reply of the communication server (b)

When the client or the equipment server is being connected, the user or the administrator must set the password for the decryption of both the public and the secret keys stored in the files.

When connecting to the communication server, the client transmits an encrypted message, which contains the working mode, the identifier, the timestamp, and the random sequence digitally signed by client (see figure 3).

The encryption is performed by means of the public key of the communication server. The communication server decrypts the packet and verifies the digital signature. If verification and timestamp check was successful, the requested working mode is set. If the public mode has been set, all further interactions are carried out without any means of cryptographic security. In the protected and the private mode the communication server acknowledges successful transition and sends the secret key (session or temporary).

The reply of the communication server is encrypted by the public key of the client and digitally signed by the communication server. The sizes of initial messages are equal.

In the protected mode, all requests from the client to the communication server are transmitted with the MAC-code; that is a 16-byte code is added to a normal request. The communication server checks the MAC-code using the secret key given to this client (figure 4). Then the communication server forwards the client's request to the equipment server. The MAC-code is calculated and added to the request; at that the secret key given to the equipment server is used. The equipment server defines the working mode dynamically. The request is checked by the protected mode; in case there is an error, it is checked by the private mode. The successful checkout allows to define the mode (when real work is done by the communication server and the equipment server, the administrator cuts off the public mode for the sake of security). The equipment server does the same, but in the reverse order.

In the private mode, all messages between the client, the communication server, and the equipment servers are transmitted encrypted. In this case a 16-byte digest is added to the initial message, and the whole request is encrypted by Rijndael algorithm (see figure 5).

---

## Administration of the System

---

The administrator access to the communication server is carried out with the help of a standard browser and a Web server. The Web server – communication server interaction is implemented with the help of servlet. Servlets are extension modules of Web servers with Java support. In the present work the servlet is used for the network exchange organization with the communication server according to the system protocol and the dynamic generation of HTML pages. The servlet provides the administrator both with the remote access to the system and with the very access mechanism. It is important to point out, that servlets have no graphic interface, do not depend on a hardware-software system platform, provide data protection (based on Java mechanisms) and simplify access to the latter. The standard Java extension – Java Servlet API is used for the servlet development.

The administrator is connected to the communication server as an ordinary client, but with the password in the "data" field of the inquiry frame and the password length indication in the "key" field, the command code is transferred in the "function" field. After the password has been verified the client receives a CID, equal to zero, which authorizes the work with additional functions, such as viewing the information on clients and allocated resources, client removal from the system, and resource deallocation. It should be noted, that the communication server does not start up a separate thread for the administrator, and serves him/her directly in StartCServer class. Thus, administrator commands are carried out in the basic thread that allows controlling affiliated threads serving clients operation.

Upon the administrator connection to the communication server and at each servlet reload the communication server sends the administrator the information contained in MainClientInfo and MainServerInfo classes. These classes contain methods for packing and transferring the information on clients, allocated resources and equipment servers. This information, in particular, includes a resource address, an owner client IP address, a unique owner client identification number (CID), an equipment server IP address, port number, on which the equipment server expects the connection with the client, and its alias. The information, received by the administrator, is shown as a table with a client work form. The form, for example, allows sending commands of all clients' disconnection along with deallocation, working client suspension and the subsequent resume, deallocation of the occupied by the client resource. After the command has been run, the information on clients and servers on administrator HTML page is updated.

Let's consider the developed software in more detail. The remote monitoring and client management of the distributed system is carried out with the help of servlet admin.java. On the servlet the doGet method is redefined. It basically provides the administrator as a client connection to the communication server, transfers commands to the server, receives the current information from it and forms the HTML page to transfer it to the administrator.

The information, including a client alias, is packed in the extended answer frame field before being sent to the administrator, namely, a unit is placed into the "data type" field, the amount of the additional bytes is placed into the "data" field. The data is followed by the main frame.

The contained in the HTML page form allows the administrator to send a set of commands to the communication server: INFO, STOP, SUSPEND, RESUME and RESOURCE FREE (correspondingly, the client and equipment server information transfer with the help of admincl and adminserv methods of MainClientInfo and MainServerInfo classes; the client removal from the communication server, and the all resource deallocation with the help of a change method of MainClientInfo class; the working client suspension and the subsequent resumption with the help of thread.suspend() and thread.resume() methods, and also the deallocation of the occupied by the client resource with the help of a change method MainClientInfo class). The HTML page is updated upon the process completion. The Thread stop(), suspend() and resume() class methods are successfully applied for work with client treads because the communication server does not initiate a separate tread for the administrator, and serves him/her in the basic communication server StartCServer class.

For the software development the Java 2 Platform Standard Edition v1.4 and Java Servlet Development Kit 2.1 packages were used. Apache server was used as a Web server.

---

## Conclusion

---

The distinct feature of the presented distributed information measurement system is that it allows combining different device interfaces along with their control computers into uniform network functioning on the basis of TCP/IP.

One of the advantages of the presented system is that the software controlling the experiment is running not at a remote computer (like in most web-technologies) but on the client's computer connected to the system through the global net. Besides, the structure of the communication server provides simultaneous access of few users to the research complexes or its subsystems.

We would like to point out that the cryptographic means of security assure integrity, non-repudiation, and confidentiality of data in the multi-user environment of the distributed information measurement system.

All the advantages of the presented structure are especially distinct while using the distributed system for education purposes. First, the simplified procedure of the equipment server creation provides easy introduction of scientific research equipment in educational process. Second, since all applications are run on the user's computer, the equipment server takes less resource, which allows to use the already connected to the equipment computer as a server without any additional requirements to its resources. Third, it is possible to organize for students not only laboratory practice tasks within limited algorithms, but also carrying out unique scientific research experiments.

As a conclusion, it is necessary to point out that the developed distributed information measurement system is used for the beam and plasma object analysis with the help of optical spectroscopy methods. In particular, the researches on excitation processes at atom-atom collisions with inert gas atoms' participation are carried out with its help [Kurskov et al, 2003] as well as the laboratory works with senior students of physical engineering department of Petrozavodsk State University.

---

## Bibliography

---

- [Zhiganov et al, 2000] E.D. Zhiganov, S.A. Kiprushkin, S.Yu Kurskov, A.D. Khakhaev. CAMAC Server for Remote Access to Physical Equipment. In: Learning and Teaching Science and Mathematics in Secondary and Higher Education. Joensuu University Press, Joensuu, 2000, pp. 170-173.
- [Gavrilov et al, 2003] S.E. Gavrilov, S.A. Kiprushkin, S.Yu. Kurskov, A.D. Khakhaev. Information system with remote access for education and research in physics. In: Mathematics and Science Education in the North-East Europe: History, Traditions & Contemporary Issues. Joensuu University Press, Joensuu, 2003, pp. 335-339.
- [Kurskov et al, 2003] S.Yu. Kurskov, A.D. Khakhaev. Excitation of Ar I atoms into  $3p5np$  states ( $4 \leq n \leq 6$ ) in binary Ar-Ar collisions. In: Northern optics 2003. Helsinki University of Technology, Helsinki, 2003, p. P012.

---

## Authors' Information

---

**Sergey Kiprushkin** – Petrozavodsk State University, Lenin St., 33, Petrozavodsk – 185910, Russia;  
e-mail: [skipr@dfc3300.karelia.ru](mailto:skipr@dfc3300.karelia.ru)

**Nikolay Korolev** – Petrozavodsk State University, Lenin St., 33, Petrozavodsk – 185910, Russia;  
e-mail: [korona@sampo.ru](mailto:korona@sampo.ru)

**Sergey Kurskov** – Petrozavodsk State University, Lenin St., 33, Petrozavodsk – 185910, Russia;  
e-mail: [kurskov@psu.karelia.ru](mailto:kurskov@psu.karelia.ru)

**Natalia Nosovich** – Petrozavodsk State University, Lenin St., 33, Petrozavodsk – 185910, Russia;  
e-mail: [natana@hotmail.ru](mailto:natana@hotmail.ru)



---

---

## International Workshop "Business Informatics"

---

---

### LEVELS OF BUSINESS STRUCTURES REPRESENTATION

**Katalina Grigorova, Plamenka Hristova, Galina Atanasova, Jennifer Q. Trelewicz**

**Abstract:** *With the increasingly large number of software frameworks available to facilitate "business modeling", it is important to understand the implications of the level of abstraction provided by the frameworks. In this paper, we discuss three levels of abstraction of framework for business modeling, including the most typical and widely accepted representatives of each level. We show that XML is an emerging standard of data interchange and integration for business modeling. We discuss these frameworks in the context of database representation, which is important for storage and retrieval of models.*

**Keywords:** *Business Modeling, Modeling Languages, Modeling Techniques, XML.*

---

#### Introduction

---

Competitive pressure, globalization, and the wide availability of Internet have made necessary the formal design of businesses. While in the past, business practices – rules, routines, procedures, and processes – could evolve in a piecemeal, isolated, and historical way, today, a rigorous and systemic design of such practices is needed, to ensure that customers' requests for products and services are processed at the satisfactory speed. Today's business modeling aims at the integration of the partial models that represent particular views on an enterprise. This means not only those models of distinctive and important parts of the enterprise should be created, but also that semantic relationships between partial models can be expressed.

The basic idea of business modeling is to offer different views on the business. The views should complement each other and thereby support a better understanding of complex systems by emphasizing appropriate abstractions. Therefore, a corresponding modeling language is used based on specific terminology that is common within particular view. It provides intuitive concepts to structure the problem domain in a meaningful way.

The baseline views should be flexible in the sense that they can be applied to any business area. Then business modeling may provide concepts that can be reused and adapted in a convenient way to detailed models for businesses in a specific market. Specialized modeling languages are one example. Other examples include reference models for certain types of industry.

Business modeling is performed on different levels of detail according to the needs of designers. Sometimes it is sufficient to create a common picture of the enterprise, while other times the use of detailed concept is required. For this reason the business modeling methods should allow various levels of abstraction:

- The highest level in the hierarchy of abstraction is related to external description allowing the users to express their view of how a given business structure looks, a type of meta-modeling language.

Often, communication between people that belong to different professional communities will not require a high level of detail. Instead it is sufficient, and helpful, to seek a common understanding of the "big picture". On the other hand, there are also specific tasks, like the re-design of a business process or the design of an object model that require the use of detailed concepts. This confirms the existence of variety of modeling languages and techniques. They are used to create the meta-model of a given business structure.

- The next level in the hierarchy of abstraction considers internal representation of business structures, which is a kind of application level.

This level of representation corresponds to the requirements of a given application. The variety of modeling languages and techniques does not demand the multiplicity of internal descriptions. It is preferable to use some standard approaches in order to allow the successful exchange of models between different environments.

Lately the emergence of XML (eXtensible Markup Language) is a first step to solve the problem of the variety modeling languages. XML is now widely accepted and acknowledged a standard. XML allows representing information in a simple, readable format, easily parsed by software tools.

- The third, and final, level of the hierarchy of abstraction is related to the real (existing) storage.

The representation on this level may be viewed in some sense as on-line database. In some implementations, the user is not allowed to access it and its format may not even be known by the user. In other implementations, it might be useful to enable the system analysts to operate directly with database models. Additionally, the effective database design leads to successful performance.

In this paper we discuss both external and internal levels of business structures representation and indicate the corresponding database models as a proper subject of efficient analysis. Sometimes we refer to modeling techniques suitable mostly for representation of business processes as the most important part of business modeling.

---

### **External Representation of Business Structures**

---

There is a large variety of meta-languages commonly used for the representation of the highest level of business structures. Generally these meta-languages involve different graphical primitives for describing the objects and connections between them.

The most significant aspect of all of the meta-languages is the importance that the notation plays in any model – it is the glue that holds the process together. Notation has three roles:

- It serves as the language for communicating decisions that are not obvious or cannot be inferred from the core itself;
- It provides semantics that are rich enough to capture all important strategic and tactical decisions;
- It offers a form concrete enough for humans to reason and for tools to manipulate.

Here we will discuss some typical and wide spread representatives of modeling tools, used for external business structures description.

#### **Flowcharting**

Flowcharting is among the first graphical modeling techniques, dating back to the 1960s. The advantages of flowcharts centre on their ability to show the overall structure of a system, to trace the flow of information and work, to depict the physical media on which data are input, output and stored, and to highlight key processing and decision points [Schriber, 1969] [Jones, 1986].

Flowcharting was initially intended to provide computer program logic representation, but, because of its flexible nature, it has been used in many other application areas as well, including business modeling. Despite its advantages, namely familiarity and ease of use, flowcharting is no longer a dominant modeling technique because it can provide only basic facilities in representing processes. Therefore, in the area of business modeling, flowcharts nowadays are typically used primarily as a simple, graphic means of communication, intended to support narrative descriptions of processes, when the latter become complicated and difficult to follow.

#### **Data Flow Diagrams**

Data Flow Diagramming (DFD) is a technique for graphically depicting the flow of data amongst external entities, internal processing steps, and data storage elements in a business process. DFDs are used to document systems by focusing on the flow of data into, around, and outside the system boundaries. In that respect, DFDs are comparable to flowcharts, differing from them basically in the focus of analysis: DFDs focus on data, instead of activities and control [Yourdon, 1989].

DFDs have been widely used for data modeling purposes and have become an ad-hoc standard notation for traditional systems analysis and design.

DFDs used to model the system's data processing and the flow of information from one process to another. They are an intrinsic part of many analysis methods. They show the sequence of processing steps traversed by the data. Each step documents an action taken to transform or distribute the data. DFDs are easy to read, making it possible for domain experts to create or to validate the diagrams [Sommerville, 2003].

### **Entity-Relationship Diagrams**

Entity-Relationship (ER) diagrams [Yourdon, 1989] are another widely used data modeling technique. ER diagrams are network models that describe the stored data layout of a system. ER diagrams focus on modeling the data present in a system and their inter-relationships in a manner that is entirely independent of the processing that may take place on that data. Such separation of data and operations may be desirable in cases where the data and their inter-relationships are complex enough to necessitate such an approach.

For the purposes of business process modeling, ER diagrams share similar limitations with DFDs. More specifically, ER diagrams focus primarily on data and their inter-relationships and hence do not provide constructs for modeling other process elements. Even more importantly, ER diagrams, unlike DFDs, do not provide any information about the functions depicted that create or use these data. Finally, ER diagrams are entirely static representations, not providing any time-related information that could drive analysis and measurement.

### **State-Transition Diagramming**

State-Transition (ST) diagrams originate from the analysis and design of real-time systems. ST diagrams attempt to overcome the limitations arising from the static nature of DFDs and ER diagrams by providing explicit information about the time-related sequence of events within a system. The notation being used by standard ST diagrams is very simple, consisting only of rectangular boxes that represent states and arrows that represent changes of state (transitions) [Quatrany, 2001].

Namely the possibility for a transition's depiction allows the usage of State-Transition diagrams as internal description tool. The explicit description of time-related sequence of data changes points out context relationship, which is in the base of the internal description.

### **Role Activity Diagramming**

Role Activity Diagrams (RADs) uses diagrammatic notation that concentrates on modeling individual or group roles within a process, their component activities and their interactions, together with external events and the logic that determines what activities are carried out and when [Huckvale, 1995]. RADs differ from most other process diagrammatic notations in that they adopt the role, as opposed to the activity, as their primary unit of analysis in process models. Due to this focus, they are mostly suitable for organizational contexts in which the human element is the critical organizational resource that process change aims to address.

### **Business Process Modeling Notation**

The Business Process Modeling Notation (BPMN) is the new standard for modeling business processes and web service process, as put forth by the Business Process Management Initiative (BPMI). BPMN is a core enabler of a new initiative in the Enterprise Architecture world called Business Process Management.

BPMN is only one of three specifications that the BMNI has developed – the other two are a Business Process Modeling Language (BPML) and a Business Process Query Language (BPQL).

BPMN specification provides a graphical notation for expressing business processes in a Business Process Diagram (BPD). The objective of BPMN is to support business process management by both technical users and business users by providing a notation that is intuitive to business users yet able to represent complex process semantics [Owen, 2003] [Stephen, 2001].

A BPD is made up of a set of graphical elements. These elements enable the development of simple diagrams that are intended to look familiar to most business analysts, resembling a flowchart-type diagram. The elements were chosen to be distinguishable from each other and to utilize shapes that are familiar to most modelers. For example, activities are rectangles and decisions are diamonds. It should be emphasized that one of the drivers for the development of BPMN is to create a simple mechanism for creating business process models, while at the same time being able to handle the complexity inherent to business processes. The approach taken to handle

these two conflicting requirements is to organize the graphical aspects of the notation into specific categories. This approach provides a small set of notation categories, easier recognition of the basic types of elements and understanding of the diagram. Within the basic categories of elements, additional variation and information can be added to support the requirements for complexity without dramatically changing the basic look-and-feel of the diagram.

---

### Internal Representation of Business Structures

---

Business structure models must be capable of providing various information elements to its users. Such elements include, for example, what activities are carried out, who is performing these activities, when and where are these activities performed, how and why are they executed, and what data elements they manipulate. Modeling techniques differ in the extent to which their constructs highlight the information that answers these questions. To provide this information, a modeling technique should be capable of representing one or more of the following “perspectives” [Curtis, 1992]:

- Functional perspective: Represents what activities are being performed.
- Behavioral perspective: Represents when activities are performed (for example, sequencing), as well as aspects of how they are performed through feedback loops, iteration, decision-making conditions, entry and exit criteria, and so on.
- Organizational perspective: Represents where and by whom activities are performed, the physical communication mechanisms used for transfer of entities, and the physical media and locations used for storing entities.
- Informational perspective: Represents the informational entities (data) produced or manipulated and their relationships.

There are known some techniques and specific languages for internal description of business structures. All of the techniques discussed here possess the above perspectives. The main benefit of one of these techniques is to build a model of a business structure that is suitable for future data processing. We discuss standard approaches here, since it is appropriate to use some standard approach to ensure interoperability between environments.

It is important to point out that the internal models describe the context data dependence, the internal relationship between processes and subprocesses and the data flow in the scope of the presenting business structure.

The presentation of some wide spread techniques for business structure representation follows.

#### IDEF Techniques

The IDEF family of modeling techniques was developed as a set of notational formalisms for representing and modeling process and data structures in an integrated fashion. The IDEF suite consists of a number of independent techniques, the most well known being IDEF0 (Function Modeling), IDEF1x (Data Modeling), and IDEF3 (Process Description Capture).

The IDEF0 method is designed to model the decisions, actions, and activities of an organization or other system and, as such, it is targeted mostly towards the functional modeling perspective (Mayer). As a communication tool, IDEF0 aims at enhanced domain expert involvement and consensus decision-making through simplified graphical devices. Perhaps the main strength of IDEF0 is its simplicity, as it uses only one notational construct, called the ICOM (Input-Control-Output-Mechanism). IDEF0 supports process modeling by progressively decomposing higher-level ICOMs into more detailed models that depict the hierarchical decomposition of activities.

Despite its advantages, IDEF0 presents a number of limitations that may render the technique unsuitable for process analysis. More specifically, IDEF0 models are static diagrams with no explicit or even implicit representation of time. Even the sequence of ICOMs is not meant to depict the temporal relations between activities. As such, IDEF0 models cannot represent the behavioral or informational modeling perspectives. To overcome some of the limitations of IDEF0 models, IDEF3 has been developed. IDEF3 describes processes as ordered sequences of events or activities. As such, IDEF3 is a scenario-driven process flow modeling technique, based on the direct capture of precedence and causality relations between situations and events. The goal of an IDEF3 model is to provide a structured method for expressing the domain experts' knowledge about how a



particular system or organization works (as opposed to IDEF0, which is mainly concerned with what activities the organization performs).

IDEF1x was designed as a technique for modeling and analysis of data structures for the establishment of database requirements. IDEF1x differs from traditional data modeling techniques in that it does not restrict the model to the data elements that are being manipulated by computers, but allows the modeling of manually-handled data elements as well. IDEF1x utilizes simple graphical conventions to express sets of rules and relationships between entity classes in a fashion similar to Entity-Relationship diagrams.

The power of IDEF1x diagrams for integrated databases can be harnessed when these diagrams are combined with IDEF0 and IDEF3 business models. Since they belong to the same "family" of techniques, IDEF models can complement each other effectively and, when combined, can provide a holistic perspective of a modeled system. However, this facility comes at a potentially high complexity of developing and maintaining many different models for a single system.

### **Petri Nets**

Petri Nets do not provide a business process structure representing technique, since they have originated from and have been traditionally used for systems modeling. However, among the systems modeling techniques, Petri Nets is perhaps the one technique that has received the most attention as a potential candidate for business process structure representing as well [Reising, 1992]. Basic Petri Nets are mathematical-graphical representations of systems, intended for assisting analysis of the structure and dynamic behavior of modeled systems, especially systems with interacting concurrent components [Peterson, 1981]. A basic Petri Net graph is composed of a set of states and a set of transitions.

It has been recognized that basic Petri Nets are not succinct and manageable enough to be useful in modeling and representing high-level, complex business processes structures. To this end, a number of extensions to the basic Petri Net formalism (usually to include the notions of "colour", "time", and "hierarchy") have been proposed [Jensen, 1996]. These extensions are collectively referred to as "high-level Petri Nets" and include, for example, Generalised Stochastic Petri Nets (GSPN), Coloured Petri Nets (CPN), and others.

The power of Petri Nets for internal description is the well-composed formalism consisting of the set of states and the state of transitions. Those familiar with this formalism can use the internal structure description for different intentions in any chosen language without environmental dependence.

### **Unified Modeling Language (UML)**

Introduced in 1997 and supported by major industry-leading companies, the Unified Modeling Language (UML) has rapidly been accepted throughout the object-technology community as the standard graphical language for specifying, constructing, visualizing, and documenting software intensive systems [Booch, 1999]. UML utilizes a wide array of diagrammatic notations, including:

- Use case diagrams, which capture system functionality as seen by the users
- Class diagrams, which capture the vocabulary of the system
- Behavior diagrams (for example state chart, activity and interaction diagrams)
- Implementation diagrams (for example, component and deployment diagrams)

The underlying reason for the development of the language is simple: although a wide variety of notational languages have long existed for the representation of software systems, most languages are typically aligned with a particular analysis and design method. This wide variety can be a source of complexity and problems of non-compatibility between languages. UML attempts to address this gap by being a "universal" language, covering everything from business process representation to database schema depiction and software components modeling. According to its developers, UML "will reduce the degree of confusion within the industry surrounding modeling languages. Its adoption would settle unproductive arguments about method notations and model interchange mechanisms, and would allow the industry to focus on higher leverage, more productive activities" [UML, 1997].

As far as business structure representation and database modeling are concerned, UML is mostly targeted to systems modeling situations, although an "extension for business structure modeling" has also been developed.

Some authors [Trelewicz, 2004] argue that UML is not appropriate for these applications because of its lack of context and structural complexity and resistance to the natural evolution of business structures. Furthermore, some may argue that the language is heavily based on the object-oriented paradigm and hold out very good possibilities for internal system representation without program environment and language dependence. There is no reason to be used in situations where the modelers want to follow only the system overview.

### **SADT (Structured Analysis and Design Technique)**

An SADT model is a simple representation one aspect of business structure, which is adequate for a given purpose. To achieve the benefits of the principles of structuring, especially top-down, levels of detail and hierarchy the model should be graphic. Each SADT model consists of a set of related diagrams, which are organized in a top-down manner. Each diagram is either a summary (parent) diagram or a detailed (child) diagram of the parent [Vernadat, 1996]. There exist two types of SADT models. An Activity model is oriented toward the decomposition of activities whereas a Data model is oriented toward the decomposition of data. Each type of model contains both activities and data; the difference lies in the primary focus of the decomposition.

A SADT Activity Model is used to describe the decomposition of activities. Data is included in the activity model as inputs, outputs, controls, and mechanisms. The top-level diagram is detailed on separate diagrams. All data is related to a given activity is explicitly shown (usually in more detail) on the lower level diagrams.

When one develops a Data model, it is not a mirror image of the Activity model, but rather the Data model is used like a data dictionary to provide a more rigorous definition of data. Experience has shown that the development of an Activity model in and by itself does not force a precise decomposition of data.

The SADT usage for internal structure representation provides the ability for explicit depiction of data and process relationships without program language and environment dependence. Each structure represented with this technique may be used in different ways and for different purposes even after its storage.

---

### **Extensible Markup Language**

Many applications use business descriptions. The problem is that these applications work with descriptions in their own internal representations. Therefore communication between them, a growing need for industry, is nearly impossible without some kind of translator. That is why some sort of exchange standard is needed in order to avoid a point-to-point translator for every pair of applications.

Extensible Markup Language (XML) is widely recognized as a rapidly emerging standard for moving data over the Internet. It is a scripting language for representing structured data in a text file. The structured data represented by XML can be virtually anything, for example, address books, configuration parameters, spreadsheets, Web pages, financial transactions, technical drawings, and so on. XML defines a set of rules for text formats for such data. By storing data in a structured text format, XML allows the user to read the data independent of the program that produced it. XML files are easy for computers to generate and read, they are unambiguous, and they avoid common pitfalls of text data formats, such as lack of extensibility, lack of support for internationalization and localization, and platform dependency.

XML offers many advantages as a general-purpose mechanism for representing data and communicating between applications [XML, 2002]:

- Flexibility

XML can be used for an enormous variety of different purposes just by defining element names and arrangements appropriate for the particular purpose. Since each element is clearly marked with begin and end tags, elements can grow and shrink as needed. Finally, the nesting property of XML elements makes it easy to combine smaller documents into larger documents.

- Portability

XML documents can be moved easily among machines or over the Internet because they are based on text, not binary representations. The text form used in XML is Unicode, which supports all of the world's languages, allowing the use of XML for applications that span national boundaries.

- Self-describing

Each XML document carries a structural description of its contents in the form of the element tags. This makes it much easier for one application to use an XML document created by a different application.

- General-purpose tools

Since all XML documents have the same basic form, general-purpose tools can be created that operate on any XML document, such as tools that create documents, display their contents, modify their structure, or record statistics about the flow of XML documents in a system. Several general-purpose XML parsers have already been created, which makes it easy to XML-enable applications.

- Robustness

Because XML documents are self-describing, XML-based applications can be built to tolerate errors and to evolve with changes in the content structure. The tags also allow graceful evolution of XML-based software. A new element can be added to a document without affecting existing software that uses the document: old software will simply ignore the new element.

- Human-readability

Although XML is intended for processing by computer programs, its textual form is also possible for humans to read. This can be useful when debugging XML-based applications and means that, if needed, a human can use an ordinary text editor to create or repair XML documents.

A software module called an XML processor is used to read XML documents and to provide access to their content and structure. It is assumed that an XML processor is doing its work on behalf of another module, called the application. This specification describes the required behavior of an XML processor in terms of how it must read XML data and the information it must provide to the application.

XML's real impact is in the area of data interchange and integration. Over the next few years XML promises to revolutionize the way that applications and enterprises exchange information. XML makes it much easier for applications to work together, even when they are in different organizations.

---

### **Database Representation of Business Structures**

---

Database models of business structures are usually hidden and the user does not access them directly, but rather through an interface on the software tools for capturing and analyzing the models. We include in our definition of "database representations" those structures that business modeling tools may utilize for file storage or analysis; i.e., we do not restrict to tables in recognized relational database middleware. In many cases the database design plays an important role for making possible the kinds of analysis that the business designer wishes to perform on the model. If the designer is authorized to have knowledge of the structure of the business model in the database, more efficient understanding, implementation and support may be possible. For example, this can allow the designer to structure the representation in such a way to facilitate faster analysis or higher degree of reuse of business process templates.

When the business designer needs to deal with only a small part of business model working with its database model is very convenient. There are a number of aspects to the enterprise, each giving a different view on it, presented by corresponding business structure.

When designing the business, the business designer will often create one aspect of the enterprise at a time, linking the aspects of the enterprise once each aspect is understood. A business designer interested in one aspect of the business can look just at that aspect of the business, without considering the others. However, having in mind the interdependence of the aspects and the affect they may have on each other, the business analyst have to take care about the connections between the different aspects.

Formally the business structures that we discuss are mostly graphs, comprising nodes and edges. Thus, the database approach is addressed to database representation of graphs. The database model must allow modifications with computational efficiency. Also, the database must support the structuring of queries appropriate for the business model. The database representation is scaleable, the memory usage associated with the business model increases linearly in respect with the of the business structures.

---

## Conclusions and Future Works

---

Our current work addresses the co-representation of business structures using XML and the BPMN standards. We have chosen these representations of the suite of prior art discussed above, leveraging the wide acceptance and ad-hoc standardization that is provided by these options.

In spite of its advantages XML has three disadvantages, all of which are inevitable consequences of XML's flexibility [Bouret, 2002]:

- Size

XML documents occupy a lot of space due to the use of text for everything and the presence of the tags. Thus, XML documents will take more space on disk and may also take more time when transmitting over a network.

- Performance

It takes time to read and write XML documents. The tags must be read and processed, and information such as numbers will have to be converted from its textual form to the form that the application needs.

- Complexity

Reading an XML document is very complicated due to the tag processing that must occur.

Despite these limitations, our subsequent work will discuss the analysis of business models utilizing XML and leveraging the large number of tools available for its creation and editing, such as Microsoft Visio or Rational Software Modeler.

---

## Acknowledgements

---

The paper presents results of research project "Database representation of business architectures for efficient analysis and modification" supported by IBM.

---

## Bibliography

---

- [Booch, 1999] G. Booch, J. Rumbaugh, I. Jacobson, Unified Modeling Language User Guide, Addison-Wesley, Reading, MA, 1999
- [Bouret, 2002] R. Bourret, XML and Databases, (<http://www.rpbouret.com>), 2002
- [Curtis, 1992] W. Curtis, M. I. Kellner, J. Over, Process Modeling, Communications of the ACM, 35, 9, 1992, pp. 75-90.
- [Huckvale, 1995] T. Huckvale, M. Ould, Process Modeling – Who, What and How: Role Activity Diagramming. In Grover, V. and Kettinger, W.J. (Eds.), Business Process Change: Concepts, Methods and Technologies, Idea Group Publishing, Harrisburg, PA, 1995, pp. 330- 349
- [IDEF, 2003] IDEF Family of Methods, A Structured Approach to Enterprise Modeling and Analysis, Knowledge Based Systems, Inc., <http://www.idef.com/>, 2003
- [Jensen, 1996] K. Jensen, Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use, Springer Verlag, Berlin, 1996
- [Jones, 1986] J. L. Jones, Structured Programming Logic: A Flowcharting Approach, Prentice Hall, New Jersey, 1986
- [Owen, 2003] M. Owen, J. Raj, BPMN and BPM. Introductions to the New Business Process Modeling Standard, PopkinSoftware, [www.popkin.com](http://www.popkin.com), 2003,
- [Peterson, 1981] J. L. Peterson, Petri Net Theory and the Modeling of Systems, Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [Quatrani, 2001] T. Quatrani, Visual Modeling with Rational Rose 2000 and UML, Addison-Wesley Publishing Company, 2001
- [Reising, 1992] W. Reising, S. S. Muchnick, P. Schnupp, (Eds.) A Primer in Petri Net Design, Springer Verlag, Berlin, 1992.
- [Sommerville, 2003] I. Sommerville, Software Engineering, Addison-Wesley Publishing Company, 2003
- [Schriber, 1969] T. J. Schriber, Fundamentals of Flowcharting, Wiley, New York, 1967
- [Stephen, 2001] A. Stephen, Introduction to BPMN, IBM Corporation, <http://www.bpmn.org/Documents/IntroductionBPMN.pdf>, 2001
- [Trelewicz, 2004] J. Q. Trelewicz, J. L. C. Sanz, D. W. McDavid, A. Chandra, S. C. Bell, Informatics for business is more than process automation: i-BUSINESS > e-PROCESS, Int'l Conf on Automatics and Informatics, SAI'04.
- [UML, 1997] UML Proposal to the Object Management Group, <http://www.rational.com/uml>, 1997
- [Vernadat, 1996] F. V. Vernadat, Enterprise Modeling and Integration: Principles and Applications, Chapman & Hall, 1996
- [XML, 2002] XML Introduction, <http://www.tcl.tk/advocacy/xmlintro.html>, 2002

---

[Yourdon, 1989] E. Yourdon, Modern Structured Analysis, Prentice Hall International, Englewood Cliffs, NJ, 1989  
[Хохлова, 2004] Хохлова М. Н., Теория Эволюционного Моделирования, ФГУП ЦНИИ-то-минформ, Москва, 2004

---

### Authors' Information

---

**Katalina Grigorova** – Department of Informatics and Information Technologies, University of Rousse, 8 Studentska St., Rousse -7017, e-mail: [katya@ami.ru.acad.bg](mailto:katya@ami.ru.acad.bg)

**Plamenka Hristova** – Department of Informatics and Information Technologies, University of Rousse, 8 Studentska St., Rousse -7017, e-mail: [pamela@ami.ru.acad.bg](mailto:pamela@ami.ru.acad.bg)

**Galina Atanasova** – Department of Informatics and Information Technologies, University of Rousse, 8 Studentska St., Rousse -7017, e-mail: [gea@ami.ru.acad.bg](mailto:gea@ami.ru.acad.bg)

**Jennifer Q. Trelewicz** – IBM Research Relationship Manager, Eastern Europe, Russia, and CIS, Research Staff Member, IBM Almaden Research Center, e-mail: [trelewicz@us.ibm.com](mailto:trelewicz@us.ibm.com)

## DECISION SUPPORT SYSTEM FOR INVESTMENT PREFERENCE EVALUATION UNDER CONDITIONS OF INCOMPLETE INFORMATION

Ivan Popchev, Irina Radeva

**Abstract:** *The proposed approach for decision support system of Investment Preference evaluation enables a categorization of public companies according to their financial stability and safety. The Investment Preference (IP) herein introduced is a qualitative criterion assumed to estimate the minimal probability of bankruptcy expressed via three categories Risky, Satisfactory and Excellent. Seven bankruptcy prediction models and Bulgarian public companies accounting information were used to illustrate the possibilities of the proposed approach. The objectives in this paper are to eliminate the subjectivity of an expert decision in evaluation of the limits of the IP categories Excellent, Satisfactory and Risky and to specify the IP categories' classification functions.*

**Keywords:** *Economics, Evaluation, Decision making, Discriminant analysis, Iterative method*

---

### RESUME<sup>1</sup>

---

The proposed decision support system for Investment Preference evaluation enables a categorization of public companies according to their financial stability and safety. The concept Investment Preference (IP) is a qualitative criterion assumed to estimate the minimal probability of bankruptcy expressed via three categories Risky, Satisfactory and Excellent.

The development of this decision support system was provoked by the incomplete Bulgarian experimental data concerning bankrupted companies which not allows a clear data discrimination about companies with positive and negative financial indicator.

The following seven popular models for bankruptcy prediction estimation were used: Altman's Z – score developed in 1968 for public companies, revised for private firms Altman's Z – score developed in 1983, Altman's "Mexican" Z – score developed in 1995, Fulmer's model developed in 1984, Springate's model developed in 1978, Russian R – Model developed in 1999 and Voronov – Maximov model developed in 2000.

The objectives in this paper are to present a simple approach for decision support system which eliminates the subjectivity of an expert decision in evaluation of the limits of the Investment Preference categories Excellent, Satisfactory and Risky and to specify the IP categories' classification functions.

---

<sup>1</sup> This work is supported by the National Science Fund of the Bulgarian Ministry of Education and Science under grand No. I1305/2003

To attain these objectives discriminant analysis and a combinatorial algorithm for the IP categories' intervals Excellent, Satisfactory and Risky specification were used. The experimental sample was designed from accounting information data about forty listed on Bulgarian Stock Exchange Sofia companies, which are assumed to be stable. This impelled assuming of a qualitative concept with positive connotation expressed via three categories Excellent, Satisfactory and Risky.

The decision support system is two staged. The first stage allows input and raw accounting data processing for selected public companies and transformation of the seven bankruptcy prediction models' scores into three aggregation groups corresponding to statuses B (Bankrupt), U (Uncertain) and N (No Bankrupt).

The second stage allows specification of the quantitative intervals of categories Excellent, Satisfactory and Risky, elimination of any expert subjectivity when determining the quantitative intervals of IP categories Excellent, Satisfactory and Risky and specification of the classification and canonical functions.

The final result generated by presented approach for decision support system for IP assessment should be interpreted as a reference point in the initial stage of investment decision making process and helps to find out at which company the interest could be raised before the further analysis performance.

---

### Authors' Information

---

**Ivan Popchev, Irina Radeva** – Institute of Information Technologies, Akad Georgy Bonchev str., bl. 2, 1113 Sofia, Bulgaria; Tel.: (+359 2) 716851, Fax: (+359 2) 9434589; e-mail: [iradeva@iit.bas.bg](mailto:iradeva@iit.bas.bg)

## АВТОМАТИЗАЦИЯ ОТБОРА СТРУКТУРИРОВАННОЙ ИНФОРМАЦИИ ДЛЯ ПОДГОТОВКИ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

**Андрей Д. Данилов**

**Аннотация:** В статье рассматривается теоретическая и практическая возможность автоматизации отбора структурированной информации из баз и хранилищ предприятия для подготовки управленческих решений. В основе теоретического подхода лежит закон причинно-следственной связи, который используется для определения наличия фундаментального смысла в структурированной информации об объекте. Его наличие является критерием для использования такой структурированной информации в задачах подготовки и принятия управленческих решений.

**Ключевые слова:** Информационная структура объекта, как его определенная информационная модель; Фундаментальный смысл, содержащейся в структурированной информации об объекте; Закон причинно-следственной связи; Автоматизация отбора структурированной информации из баз и хранилищ; Хранилище структурированной информации, в которой содержится фундаментальный смысл.

---

### Введение

---

Международный и российский опыт показывает, что конкурентная способность крупных предприятий существенно зависит от эффективности стратегического и текущего планирования и управления. Основу этих процессов составляют решения трех важнейших задач, а именно:

- **определение** стратегической и текущей **целевой функции развития и управления** предприятием;
- **решение задачи анализа** текущей ситуации на предприятии, рынках потребления его продукции, возможностей субподрядчиков, поставщиков, конкурентов и т.д.;
- **подготовка** стратегического и текущих **планов** работы предприятия и **управленческих решений** для их реализации.

Соответственно на многих предприятиях для подготовки решения этих задач созданы информационно-аналитические подразделения, которые в своей работе используют различные базы и хранилища открытой и корпоративной информации, а также специализированные информационные, аналитические и экспертные системы.

Эффективность решения этих задач напрямую связана с наличием необходимой и достоверной информации в базах и хранилищах предприятия. Руководители и аналитики информационно – аналитических и ситуационных центров предприятий хорошо знают, что “информационный мусор”, который не имеет содержательного смысла, весьма затрудняет исследовательскую работу и увеличивает время выполнения анализа для подготовки управленческих решений.

Если эти специалисты будут иметь **инструмент** для **автоматического** отсеивания информации, которая не несет содержательного смысла для решения конкретной задачи, то время анализа будет использовано на получение и обработку только той информации, которая может быть полезна для подготовки соответствующих рекомендаций и управленческих решений. С нашей точки зрения такой информацией является структурированная информация об объектах управления, которая формируется на основе информационной модели объекта в виде его определенной информационной структуры.

В этой статье мы обсудим возможность **идентификации** наличия фундаментального **смысла** и его **количества** в структурированной информации об объектах, и приведем алгоритмы для ее технической реализации. Решение этой задачи дает возможность **автоматизировать** процесс отбора структурированной информации, необходимой для задач анализа, подготовки рекомендаций и управленческих решений в информационно-аналитических подразделениях предприятия.

---

### **Количество смысла в структурированной информации об объекте**

---

Одним из важных направлений теории информации является решение задачи автоматической **идентификации смысла**, заключенного в информации. Можно предположить, что в результате ее решения будет получена фундаментальная основа для создания **искусственного интеллекта**.

Мы рассмотрим более частную задачу – существует ли принципиальная возможность использовать технические устройства для автоматического определения наличия **смысла** и его **количества** в сообщении или хранилище данных, которые содержат структурированную контекстно-независимую информацию об объектах. Ее решение является только определенным шагом на пути к решению задачи автоматической **идентификации смысла**, содержащегося в структурированной информации об объекте.

Известно, что классическую теорию передачи, получения и обработки сигналов часто называют теорией информации, которая базируется на фундаментальных работах К. Шеннона [1] и Р. Хартли [2]. В ее основе лежит **статистическое** описание источников сообщений и каналов связи, а также **статистическое** измерение **количества информации** в сообщении, которое определено только вероятностными свойствами сообщений и не от каких других их свойств не зависит.

Необходимо заметить, что данная теория не позволяет определить наличие **смысла** и его **содержания** в информации, которая находится, как в источнике, так и в приемнике сообщения.

Заметим, что базы и хранилища структурированной информации предприятия могут быть, как источником, так и приемником соответствующих информационных сообщений, относящихся к определенным объектам. Например, к таким объектам можно отнести персонал предприятия, результат его деятельности, основные фонды предприятия и т.д.

Для решения задачи идентификации количества смысла в структурированной информации, содержащейся в соответствующем источнике или в приемнике информации необходимо дать определение термина **смысл**, который заключен в структурированной информации об объектах. При этом под термином структурированная информация об объекте будем понимать определенную информационную модель объекта, которую назовем **информационной структурой объекта**.

Информационная структура объекта определенным образом, т.е. информационно характеризует и связывает **результат** изменения состояния объекта, **его способности** изменять свое состояние и **причины**, вызывающие изменение состояния объекта [3]. При этом информация, представленная

в информационной структуре объекта может иметь любую форму, например, буквенную, цифровую, графическую и т.д.

В наиболее общем виде информационную структуру объекта можно представить следующей символьной записью, как [3]

$$I_o = fo_e (le), \quad (1)$$

где  $I_o$  - информационное представление состояния объекта,  $fo_e$  - информационное представление способности объекта изменять свое состояние,  $le$  - информационное представление причины, вызывающей изменение состояния объекта.

Такую информационную структуру объекта можно назвать **информационной моделью объекта**, которая на основе наших знаний может быть наполнена соответствующей информацией, и располагаться в базах или хранилище предприятия. Далее мы покажем, что такая информационная модель объекта наиболее удобна для оценки наличия фундаментального смысла в структурированной информации об объекте. Кроме этого, такая информационная модель является определенным базисом для автоматизации процесса отбора структурированной информации, которую наиболее целесообразно использовать для анализа и подготовки управляющих решений.

Для удобства дальнейшего изложения материала обозначим через  $n = n_i$  количество символов в алфавите сообщения об объекте. Соответственно, если сообщение об объекте будет в виде его информационной структуры (см. выражение 1), то количество символов в таком сообщении равно трем, т.е.  $n_i = (n_1 = le; n_2 = fo_e; n_3 = I_o)$  при  $i = 1, 2, 3$ .

Предположим, что в отправленном или принятом сообщении вероятность появления любого  $n_i$ -ого символа не меняется с течением времени. Следовательно, такое сообщение называется «шеноновским» и среднее количество информации в таком сообщении можно определить по формуле К. Шеннона

$$H = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

где  $H$  - среднее количество информации в сообщении;  $p_i$  - вероятность появления  $i$ -ого символа алфавита.

Предположим, что в сообщении об объекте  $n_i = (n_1 = le; n_2 = fo_e; n_3 = I_o)$

при  $i = 1, 2, 3$  вероятность появления  $i$ -ого символа алфавита  $p_i = 1/3$ .

Следовательно, среднее количество информации в сообщении определяются на основе выражения 2, как

$$H = - (1/3 \log_2 1/3 + 1/3 \log_2 1/3 + 1/3 \log_2 1/3) \quad (3)$$

Выражение 2 показывает, какое количество информации необходимо для того, чтобы при передаче техническим устройством определенного сообщения можно «узнать» символы алфавита, т. е. выделить из сообщения тот или иной знак и **не более**. Предположим, что эта процедура произошла, т.е. из сообщения об объекте выделены все символы, а именно  $le; fo_e; I_o$ .

Можно ли далее определить наиболее общий, т.е. фундаментальный **смысл**, который заложен в этом сообщении?

Если проанализировать это сообщение, которое на основе **трех символов** определенного алфавита соответствующим образом представляет собой наиболее общее описание всех составляющих информационной структуры **любого объекта**, и при этом характеризуют его динамические и статические характеристики, то становится понятным, что такое сообщение содержит только **три** фундаментальных смысла, а именно:

- $I_o$  теоретически точно определяется взаимосвязью между  $fo_e$  и  $le$ ;
- $fo_e$  теоретически точно определяется взаимосвязью между  $I_o$  и  $le$ ;
- $le$  теоретически точно определяется взаимосвязью между  $fo_e$  и  $I_o$ .



Можно сказать, что эти **три** фундаментальных смысла, которые содержит сообщение в виде такого информационного представления объекта, следуют из логики, заключенной во взаимосвязях **составляющих информационной структуры** объекта, т.е.

$$\begin{array}{lll} & & \text{смысл1 - } \mathbf{foe} = \mathbf{lo(le)}, \\ \text{если } \mathbf{lo, foe, le} & \text{то} & \text{смысл2 - } \mathbf{lo} = \mathbf{foe (le)}, \\ & & \text{смысл3 - } \mathbf{le} = \mathbf{lo(foe)}. \end{array} \quad (4)$$

При этом необходимо отметить, что если количество информации (см. выражения 2,3) в сообщении является определенной **вероятностной** оценкой появления любого  $n_i$ -ого символа при ( $i = 1, 2, 3$ ), т.е. **lo, foe, le**, то количество фундаментального смысла в этом сообщении, если все символы сообщения определены, всегда **детерминировано**.

Кроме этого, можно утверждать, что количество фундаментального смысла в таком сообщении **теоретически точно** определено, т.е. такое сообщение содержит **три** фундаментальных смысла.

Таким образом, можно ввести следующее правило определения **количества** фундаментального смысла в сообщении об объекте, алфавит которого состоит из символов, характеризующих все три взаимосвязанные составляющие информационной структуры объекта.

#### Определение 1

*Если источник или приемник сообщения содержит алфавит, в котором соответствующие символы характеризуют информационное представление **состояния объекта**, информационное представление **способности объекта** изменять свое состояние и информационное представление **причины**, вызывающей изменение состояния объекта, то количество фундаментального смысла в этом сообщении **три**, и они теоретически точно определяются взаимосвязью:*

- информационного представления **способности объекта** изменять свое состояние и информационного представления **причины**, вызывающей изменение состояния объекта;
- информационного представления **состояния объекта** и информационного представления **причины**, вызывающей изменение состояния объекта;
- информационного представления **состояния объекта** и информационного представления **способности объекта** изменять свое состояние.

В этом определении заключен фундаментальный философский смысл **причинно-следственной связи**, который можно охарактеризовать следующим образом.

#### Определение 2

*Если известны **причина**, которая вызвала изменение состояния объекта, **способность** объекта изменять свое состояние в зависимости от этой причины и **следствие**, как результат изменения состояния объекта, то всегда существует точная связь в виде отношений:*

- между **следствием** и **причиной**;
- между **способностями** и **следствием**;
- между **причиной** и **способностями**.

Соответствующим образом, можно определить количество фундаментального смысла в **любом количестве** независимых источников сообщений об объектах, состоящих из алфавита, в котором имеются соответствующие символы, характеризующие информационное представление **состояния** объекта, информационное представление **способности** объекта изменять это состояние и информационное представление **причины**, вызывающее изменение состояния объекта, т.е.

$$\begin{array}{l} 3 \quad x \\ \mathbf{Kn}_{ji} = \sum_{j=1}^3 \sum_{i=1}^x \mathbf{M}_i \end{array} \quad (5)$$

при  $j = (1, 2, 3)$ ;  $i = (1, 2, 3, \dots, x)$

если  $j = 1$ , то ( $n_{i1} = lo$  или  $n_{i1} = foe$  или  $n_{i1} = le$ ),

если  $j = 2$ , то ( $n_{i1} = lo, n_{i2} = foe$ ) или ( $n_{i1} = lo, n_{i2} = lo$ ) или ( $n_{i1} = le, n_{i2} = foe$ )

если  $j = 3$ , то ( $n_{i1} = lo, n_{i2} = foe, n_{i3} = le$ )

где  $Kn_{ji}$  - количество фундаментального смысла в **любом количестве независимых источников сообщения**, состоящих из алфавита, в котором соответствующие  $n_{ji}$ -ые символы характеризуют **lo** - информационное представление состояния объекта, **foe** - информационное представление способности объекта изменять свое состояние и **le** - информационное представление причины, вызывающей изменение состояния объекта;  $M_i$  - соответственно  $i$ -ый источник сообщения,  $N_{ji}$  - соответственно  $j$ -ое количество фундаментального смысла в  $i$ -ом источнике сообщения.

Несложно показать, что количество фундаментального смысла  $Kn_{ji}$  для объединенного количества **независимых** источников сообщений, например, двух источников с соответствующим количеством смысла  $Kn_{j1}$  и  $Kn_{j2}$  можно рассматривать как один источник, одновременно реализующий количественную пару фундаментального смысла  $Kn_{j1}$  и  $Kn_{j2}$ .

Запишем следующую функцию, связывающую количество фундаментального смысла в объединенных независимых источниках сообщений при равной вероятности состояний этих источников

$$F(Kn_{j1}, Kn_{j2}) = F(Kn_{j1}) + F(Kn_{j2}) \quad (6)$$

Можно математически точно и строго показать, что единственной функцией, при перемножении аргументов которой значение функции складывается, является логарифмическая функция, которая обладает свойством адитивности, т.е.

$$-\log (Kn_{j1} \times Kn_{j2}) = -[\log (Kn_{j1}) + \log (Kn_{j2})] \quad (7)$$

Знак «-» в выражении 7 использован для того, чтобы количество фундаментального смысла в объединенных независимых источниках структурированной информации об объектах не было отрицательной величиной. При этом основание логарифмов может быть любое, например, 2. В этом случае количество фундаментального смысла в объединенных независимых источниках сообщений, содержащих структурированную информацию об объектах можно измерять в битах (от английского сочетания **binary digit**, что означает «двоичный разряд» или «двоичная цифровая единица»).

Несложно показать, что для сообщений об объектах, в которых алфавит содержит все символы, соответственно характеризующие информационное представление состояния объекта, информационное представление способности объекта изменять свое состояние и информационное представление причины, вызывающее изменение состояния объекта, количество фундаментального смысла определяется, как

$$Kn_{ji} = \sum_{j=1}^3 \sum_{i=1}^x N_{ji} M_i = \sum_{i=1}^x 3i M_i \quad (8)$$

при  $j = (1, 2, 3)$ ;  $i = (1, 2, 3, \dots, x)$  если  $j = 3$ ; ( $n_{i1} = lo, n_{i2} = foe, n_{i3} = le$ )  $\Rightarrow N=3$

или

$$Kn_{ji} = - \sum_{i=1}^x \log (3i M_i) \quad (9)$$

при  $j = (1, 2, 3)$ ;  $i = (1, 2, 3, \dots, x)$  если  $j = 3$ ; ( $n_{i1} = lo, n_{i2} = foe, n_{i3} = le$ )  $\Rightarrow N=3$

Необходимо отметить, что при этом не важно, какие символы и на каком языке представлены в сообщении для описания информационной структуры объекта. Важно, чтобы эти символы точно

определяли взаимосвязи или отношения между составляющими информационной структуры объекта, даже если они неизвестны передающей или получающей стороне.

Рассмотрим можно ли **теоретически точно** определить количество фундаментального смысла в сообщении, алфавит которого содержит символы, характеризующие любую, но **одну** из трех составляющих информационной структуры объекта.

Для решения этой задачи будем использовать определения 1 и 2, которые позволяют идентифицировать количество фундаментального смысла, содержащегося в сообщениях об информационной структуре объекта.

Предположим, что в сообщении об объекте имеется только один символ алфавита, например,  $n_1 = lo$ . Из анализа взаимосвязей составляющих информационной структуры следует, что

$$\begin{array}{ll} \text{если } n_j = n_1 = lo & \text{то} \\ \text{при } n_2 = foe = 0, n_3 = le = 0 & \end{array} \quad \begin{array}{l} \text{смысл1 - } foe = lo(le) = 0, \\ \text{смысл2 - } lo = foe(le) = 0, \\ \text{смысл3 - } le = lo(foe) = 0. \end{array} \quad (10)$$

Таким образом, из выражения 10 следует, что если в сообщении об объекте имеется только один символ алфавита, то такое сообщение **не содержит** фундаментального смысла об объекте, т.е. его количество в таком сообщении равно нулю.

Это означает, что

$$K_{n_{ji}} = \sum_{j=1}^3 \sum_{i=1}^x N_{ji} M_i = 0 \quad (11)$$

если  $j = 1$ ;  $i = (1, 2, 3, \dots, x)$ , то при ( $n_{i1} = lo$  или  $n_{i1} = foe$  или  $n_{i1} = le$ )  $\Rightarrow N_{ji} = 0$

или

$$K_{n_{ji}} = - \log \sum_{i=1}^x N_{ji} M_i = 0 \quad (12)$$

если  $j = 1$ ;  $i = (1, 2, 3, \dots, x)$ , то при ( $n_{i1} = lo$  или  $n_{i1} = foe$  или  $n_{i1} = le$ )  $\Rightarrow N_{ji} = 0$

Рассмотрим, можно ли **теоретически точно** определить количество фундаментального смысла в сообщении об объекте, которое содержит символы алфавита, характеризующие только любые две составляющие информационной структуры объекта?

Для решения этой задачи также будем использовать определения 2 и 3. Предположим, что в алфавите сообщения имеется следующая пара символов  $n_1 = lo$ ,  $n_2 = foe$ . Из анализа взаимосвязей составляющих информационной структуры объекта следует, что

$$\begin{array}{ll} \text{если } n_1 = lo, n_2 = foe & \text{то} \\ \text{при } n_3 = le = 0 & \end{array} \quad \begin{array}{l} \text{смысл1 - } foe = lo(le) = 0, \\ \text{смысл2 - } lo = foe(le) = 0, \\ \text{смысл3 - } le = lo(foe) = 0. \end{array} \quad (13)$$

Такой результат следует из **фундаментальных основ** решения обратной задачи анализа, которая принципиально не может быть решена теоретически точно **во всех случаях** (нельзя теоретически точно определить  $le$ ), например, когда **foe** содержит неоднозначную нелинейность типа гистерезис или нелинейность с зоной нечувствительности и т.д.

Таким образом, из выражения 13 следует, что если в сообщении об объекте имеется алфавит, состоящий из двух символа  $n_1 = lo$ ,  $n_2 = foe$ , то такое сообщение **не содержит фундаментального смысла** об объекте, т.е. его количество в таком сообщении равно нулю. Это означает, что

$$K_{n_{ji}} = \sum_{j=1}^3 \sum_{i=1}^x N_{ji} M_i = 0 \quad (14)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i1} = lo, n_{i2} = foe) \Rightarrow N_{ji} = 0$

или

$$K_{n_{ji}} = -\log \sum_{i=1}^x N_{ji} M_i = 0 \quad (15)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i1} = lo, n_{i2} = foe) \Rightarrow N_{ji} = 0$

В некоторых случаях, например, когда **foe** не имеет ограничений, зон нечувствительности, гистерезиса и представляет собой в частности линейную передаточную функцию, обратная задача анализа решается теоретически точно (теоретически точно определяется **le**) и количество фундаментального смысла в таком сообщении об объекте три. Такие частные случаи необходимо рассматривать отдельно при решении задачи определения количества фундаментального смысла в сообщении **о конкретном** объекте, если о нем имеется соответствующая дополнительная информация.

Поэтому руководители и специалисты информационно - аналитических подразделений предприятия должны хорошо понимать, что **не существует теоретически точного** решения обратной задачи анализа для всех возможных случаев. А это означает, что **невозможно** создать технические устройства, например, экспертные системы, которые способны точно решать такие задачи.

Предположим, что в сообщении об объекте имеется следующая пара символов алфавита  $n_1 = lo, n_3 = le$ .

Из анализа взаимосвязей составляющих информационной структуры объекта следует, что

если $n_1 = lo, n_3 = le$	то	смысл1 - $foe = lo(le) = 0,$	(16)
при $n_2 = foe = 0$		смысл2 - $lo = foe(le) = 0,$	
		смысл3 - $le = lo(foe) = 0.$	

Этот результат анализа следует из **фундаментальных основ** невозможности теоретически точного решения задачи идентификации **способности** объекта изменять свое состояние при условии, что известно только состояние объекта и причина, вызвавшая это состояние.

Такая задача анализа соответствует определению содержания **«черного ящика»**. Как известно, она принципиально не может быть решена теоретически точно **во всех случаях**. Нельзя теоретически точно определить **foe**, так как мы заранее не знаем, например, содержит ли **foe** неоднозначную нелинейность типа гистерезис или нелинейность с зоной нечувствительности и т.д. Для ее решения необходимо знать дополнительную информацию.

Таким образом, из выражения 16 следует, что если имеется сообщение об объекте, состоящее из двух символов  $n_1 = lo, n_3 = le$  алфавита, то такое сообщение **не содержит** фундаментального смысла об объекте, т.е. его количество в таком сообщении равно нулю.

Это означает, что

$$K_{n_{ji}} = \sum_{j=1}^3 \sum_{i=1}^x N_{ji} M_i = 0 \quad (17)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i1} = lo, n_{i3} = le) \Rightarrow N_{ji} = 0$

или

$$Kn_{ji} = - \log \sum_{i=1}^x N_{ji} \quad M_i = 0 \quad (18)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i1} = lo, n_{i3} = le) \Rightarrow N_{ji} = 0$

В некоторых частных случаях, например, когда мы **априорно знаем**, что **foe** не имеет ограничений, зон нечувствительности, гистерезиса и представляет собой в частности линейную передаточную функцию, задача идентификации **foe** может быть решена теоретически точно и количество фундаментального смысла в таком сообщении об объекте **три**. Такие частные случаи необходимо рассматривать отдельно при решении задачи определения количества фундаментального смысла в сообщении **о конкретном** объекте, если о нем имеется соответствующая дополнительная информация.

Предположим, что в сообщении имеется следующая пара символов алфавита  $n_2 = foe, n_3 = le$ .

Из анализа взаимосвязей составляющих информационной структуры объекта следует, что

$$\begin{array}{ll} \text{если } n_2 = foe, n_3 = le & \text{смысл1 - } foe = lo(le) = 1, \\ \text{при } n_1 = lo = 0 & \text{то } \text{смысл2 - } lo = foe(le) = 1, \\ & \text{смысл3 - } le = lo(foe) = 1. \end{array} \quad (19)$$

Такой результат определения количества фундаментального смысла в сообщении об объекте, следует из **фундаментальных основ** решения прямой задачи анализа. Она принципиально может быть решена теоретически точно **во всех случаях**, так как мы знаем **le, foe** и не имеет значения, имеются ли в **foe** какие либо нелинейности.

Заметим, что вычислительная точность решения этой задачи может зависеть от методов расчета, но для данного анализа определения количества фундаментального смысла в сообщении это не имеет существенного значения.

Из выражения 19 следует, что если сообщение об объекте имеет два символа  $n_2 = foe, n_3 = le$  алфавита, то такое сообщение **содержит** фундаментальный смысл об объекте и его количество в таком сообщении равно **трем**. Это означает, что

$$Kn_{ji} = \sum_{i=1}^x 3i \quad M_i \quad (20)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i2} = foe, n_{i3} = le) \Rightarrow N_{ji} = 3$

или

$$Kn_{ji} = - \log \sum_{i=1}^x 3i \quad M_i \quad (21)$$

если  $j = 2$ ;  $i = (1, 2, 3, \dots, x)$ , то при  $(n_{i2} = foe, n_{i3} = le) \Rightarrow N_{ji} = 3$

Таким образом, мы получили выражения (8, 9, 11, 12, 14, 15, 17, 18, 20, 21), на основе которых можно определять количество фундаментального смысла, имеющегося в любом количестве независимых сообщений об объекте. При условии, что алфавит сообщений содержит символы, характеризующие **информационное представление** состояния объекта, **информационное представление** способности объекта изменять свое состояние и **информационное представление** причины, вызывающей изменение состояния объекта.

Соответственно, эти выражения можно использовать, для создания технического средства - **счетчика** количества фундаментального смысла, содержащегося в базах и хранилищах структурированной информации об объектах предприятия.

### Счетчик количества фундаментального смысла в структурированной информации об объектах предприятия.

Можно показать, что существует **минимальный объем** алфавита сообщения об объекте, в символах которого теоретически точно содержится **все количество** фундаментального смысла об объекте. При этом предполагаем, что символы алфавита действительно соответствуют определенным составляющим информационной структуры. Из проведенного анализа (см. выражения 20, 21) следует, что таким алфавитом с минимальным объемом символов в сообщении, в котором теоретически точно содержится **все количество** фундаментального смысла об объекте, является следующий

$$a = (n_2, n_3) = \min(A) \quad (22)$$

$$A \subset a, n_2 = \text{foe}, n_3 = \text{le}.$$

где **a** - минимальный объем алфавита контекстно-независимого сообщения об объекте, в символах которого теоретически точно содержится все количество фундаментального смысла об объекте; **n<sub>2</sub>, n<sub>3</sub>** - символы алфавита, соответственно характеризующие информационное представление способностей объекта изменять свое состояние и информационное представление причины, вызывающее изменение состояние объекта; **A** - множество алфавитов контекстно-независимого сообщения об объекте, в символах которого теоретически точно содержится все количество фундаментального смысла об объекте.

Минимальность алфавита **a** сообщения об объекте, в символах которого теоретически точно содержится все количество фундаментального смысла об объекте, объясняется тем, что существует **теоретически точное решение** прямой задачи анализа.

Из проведенного анализа следует, что все остальные алфавиты контекстно-независимого сообщения об объекте являются, либо недостаточными (см. выражения 11, 12, 14, 15, 17, 18) так как не позволяют во всех случаях теоретически точно определить **количество фундаментального смысла** в этом сообщении, либо являются избыточными (см. выражения 8, 9).

Важность этого вывода заключается в том, что можно определенным образом оптимизировать объем хранилища структурированной информации об объектах, с которыми удобно работать специалистам информационно-аналитических подразделений предприятия. Это можно реализовать, если в соответствующем хранилище предприятия собирается только та структурированная информация, которая **содержит фундаментальный смысл** об объектах.

Поэтому для работы специалистов информационно-аналитических подразделений или центров предприятий целесообразно иметь устройства, которые могут автоматически выделять из общего хранилища структурированной информации только ту часть, которая содержит фундаментальный смысл об объектах.

На рис.1 представлена наиболее общая структурная схема такого технического устройства. Она содержит **приемник** сообщений в виде определенной структурированной информации об объекте, **анализатор** содержания фундаментального смысла в сообщении и **счетчик** количества фундаментального смысла в сообщении об объекте.

Из структурной схемы видно, что техническая задача по реализации анализатора решается на основе использования программных средств, в которых имеется возможность использовать классическую логику. Также понятно, что техническая реализация такого счетчика не имеет принципиальных затруднений.

Алгоритм работы такого технического устройства предполагает наличие в алфавите (**a<sub>1</sub>, ..., a<sub>7</sub>**) любого сообщения не более трех символов, т.е. **n<sub>1</sub>, n<sub>2</sub>, n<sub>3</sub>**. Такое устройство в виде соответствующего программного обеспечения наиболее просто создать, если хранилище структурированной информации предприятия реализовано на основе объектно-ориентированного подхода. Это позволяет структурированную информацию об объектах содержать в соответствующих разделах или полях хранилища предприятия, т.е.:

- информационных представлений **состояний объекта**;
- информационных представлений **способностей объекта** изменять свое состояние;
- информационных представлений **причин**, вызывающих изменение состояния объекта.

При этом существенно упрощается автоматизация идентификации количества фундаментального смысла в структурированной информации об объекте. Приемник сообщений получает сигналы о заполнении соответствующих полей хранилища предприятия, а аналитическое устройство определяет наличие в структурированной информации об объекте фундаментального смысла. Если результат положительный, то такая информация об объекте может быть эффективно использована для анализа специалистами информационно-аналитических подразделений и центров предприятий. Кроме этого, такой подход удобен для выявления информации, которая имеет определенный приоритет для системы защиты информационных ресурсов предприятия или организации.

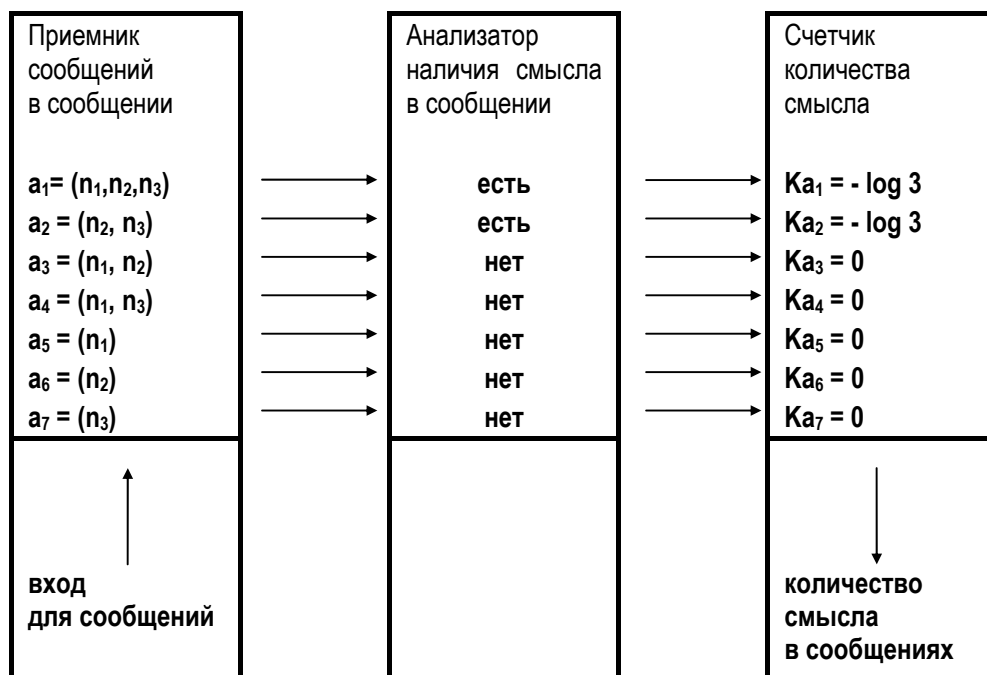


Рис.1 Структурная схема технической системы для определения количества фундаментального смысла в хранилище структурированной информации предприятия

## Литература

1. Шенон К. Математическая теория связи // К. Шенон. Работы по теории информации и кибернетики. М.: ИЛИ, 1963.- с.243-332.
2. Hartley R.V.L. Transmission of information // BSTJ.-1928.-V.7 - N3. P.535-536/
3. Вебер А.В., Данилов А.Д., Шифрин С.И. Knowledge - технологии в консалтинге и управлении предприятием - СПб.: Наука и Техника, 2003.-176с

## Информация об авторе

**Данилов А.Д.** – Санкт-Петербург, Эксперт департамента по информатизации и телекоммуникациям Ленинградской области, Эксперт некоммерческого партнерства “Учебный и исследовательский центр “Протеи”, кандидат технических наук (Phd); e-mail: [danilov@proteus-spb.ru](mailto:danilov@proteus-spb.ru), [www.proteus-spb.ru](http://www.proteus-spb.ru)

## VIABLE MODEL OF THE ENTERPRISE – A CYBERNETIC APPROACH FOR IMPLEMENTING THE INFORMATION TECHNOLOGIES IN MANAGEMENT

Todorka Kovacheva

**Abstract:** *The purpose of the current paper is to present the developed methodology of viable model based enterprise management, which is needed for modern enterprises to survive and growth in the information age century. The approach is based on Beer's viable system model and uses it as a basis of the information technology implementation and development. The enterprise is viewed as a cybernetic system which functioning is controlled from the same rules as for every living system.*

**Keywords:** *enterprise strategy, viable system model, enterprise model, neural network, artificial intelligence, cybernetics, business trends.*

---

### Introduction

---

The enterprises in the information age need to be managed in different way. The traditional management techniques successfully applied in the industrial companies are not suited in the new economy. The reason is that the conditions from the past are changed rapidly. Thus the contemporary business is accomplished in highly dynamic environment and adaptation capabilities are needed. New business trends [Kovacheva, Toshkova, 2005] have to be taken into consideration. According to this, the traditional software technologies are limited in their effectiveness, as they are unable to discover and maintain the information, which is hidden, in large amounts of data. New kind of software [Kovacheva T., 2004] is needed and new information technologies must be applied.

The main challenge for the modern enterprises is to keep their viability. To do this and because of the environment complexity and the complexity of the enterprise itself, the enterprise must be managed as a cybernetic system. Thus the suggested in this paper novel approach for enterprise management is based on cybernetics and system theory. A viable model of the enterprise is developed where the needed information technologies are applied. It is based on viable system model (VSM) [Beer S., 1984] which is the basis for our methodology.

---

### Viable System Model

---

Viable System Model is the "whole system" theory. It is developed from Stafford Beer [Beer S., 1956, 1959, 1967, 1979, 1981, 1984, 1985] who is called the father of managerial cybernetics. He was inspired from the way the human brain organizes the operation of the muscles and organs and synchronizes all the activities in human organism. VSM is a new way of thinking about organizations based on system theory and viability. Beer considers the human organism as three main interacting parts:

- muscles and organs;
- nervous system;
- external environment.

They are included in Viable System Model as follows:

1. The Operation: the units which do the basic work (muscles and organs);
2. The Metasystem: provide a service to the Operations units and ensures they work together in an integrated and harmless fashion (nervous system);
3. The Environment: all the environment elements, which are of direct relevance to the system in focus (external environment).

These three parts must be in balance. When the environment changes the enterprise must respond accordingly.

Figure 1 shows the five interacting systems in relation to the human system [Beer S., 1981]. These five systems are the basis of the Viable System Model. In Table 1, they are explained from a management point of view.



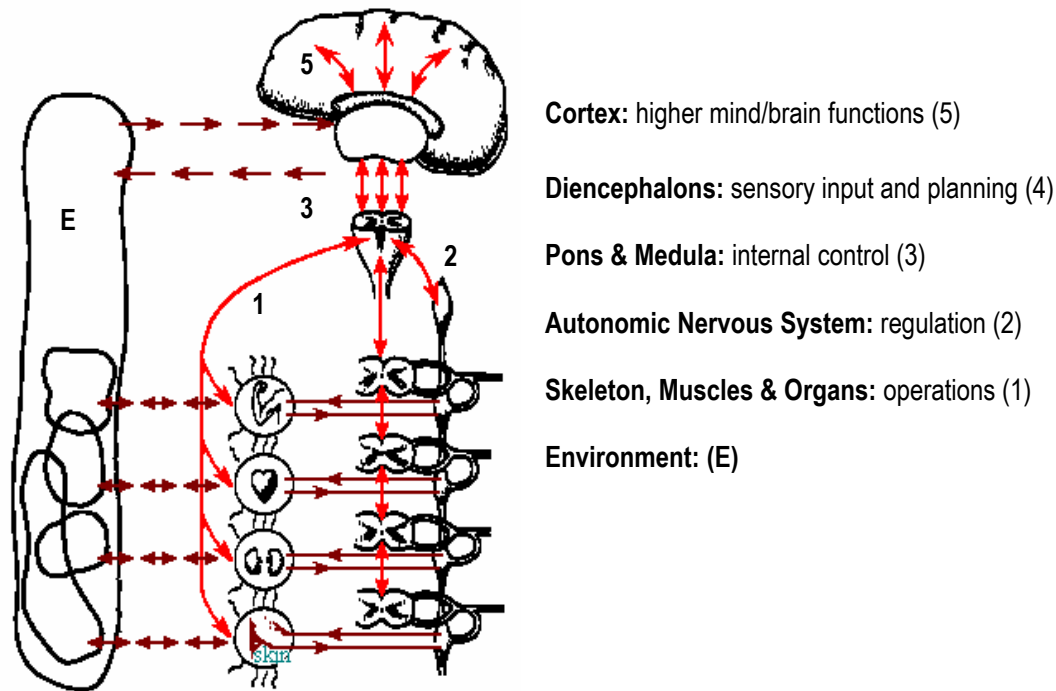


Figure 1: The Five interacting systems – neurophysiological approach

Table 1

System Number	System Identification
System 1 (S1)	Primary activities.
System 2 (S2)	Stability and conflict resolution.
System 3 (S3)	Internal regulation and optimisation.
System 4 (S4)	Sensors, adaptation, planning, strategy development.
System 5 (S5)	Policy, identity, goals

Using the Beer's Viable System Model we developed a methodology of viable model-based enterprise management where the needed information technologies for supporting business activities and keep the enterprise viable are applied.

### **Viable Model-Based Enterprise Management**

The goal of Cybernetics is to understand and formalize the basic, underlying principles of systems, such as living systems and to study the problems of complex systems, adaptation and self-organization. The main characteristic of living systems is their viability. A viable system has the capability to successfully deal with the complexity of its environment and is adaptable over time. Thus, the enterprise management must ensure that realization of the company strategy will keep it viable. Therefore, the relevant software is needed.

Operations are presented from the basic units in the enterprise. They do the actual work and could be departments, machines, people etc. according to the enterprise scale, activities and structure. These units

need to be monitored continuously to ensure they work in the proper way. Thus, the real-time software must be implemented. Such kind are of software are the well known operational systems and OTLP systems.

The daily operations in every enterprise department are registered in specific software and the department database is built. It stores all the data from the everyday activities. This information is useful for the detailed analysis of the enterprise data. This kind of software must be present in System 1.

System 2 function is to prevent and resolve conflicts. Applying the proper software the conflicts can be early recognized and prevented. The main principles of building such kind of software are given in [Kovacheva, 2004]. At this level, we need detailed and granulated data for the neural network learning process. This information than is analysed, compared and managed according specific rules, included in a conflict resolution and stability preserving expert system. The data can be organized in traditional data based as well as in data marts.

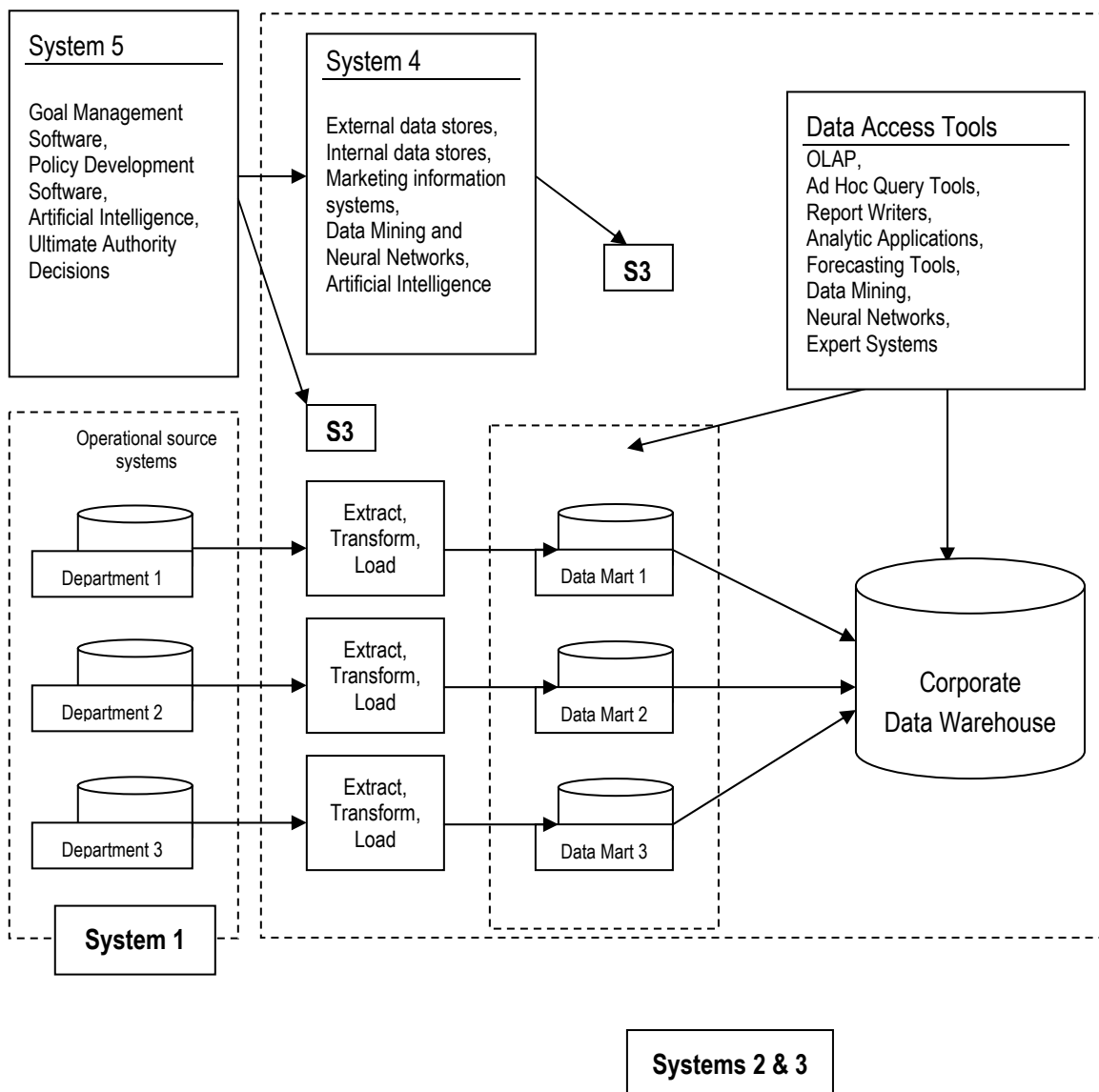


Figure 2: The Interconnections between the five systems and their elements

---

System 3 needs all the data for the enterprise everyday operations. Thus, we use the data warehousing approach, which integrates data from the operational systems into one common data store, known as data warehouse. It is optimized for data analysis purposes and decision making. We use different tools to access the data in the warehouse and discover hidden information in them.

System 4 is responsible for the enterprise adaptation. Therefore, it needs information about the external environment so it can produce strategies. It also needs a good model of the internal capabilities so it knows what tools it has at its disposal. We use different tools for forward planning and strategy development with combination of expert system, neural networks and marketing information systems (MIS). MIS [Недева В., 2003] maintain data from internal and from external sources. Both could be integrated in a big corporate data warehouse. On its basis, the external environment is analysed and the adaptation capabilities are developed.

System 5 is compared with the higher human brain functions. It is responsible for policy development, goal settings, ultimate authority and identity. At this level, a special kind of software is needed. The Artificial Intelligence is now applied. Goal definition must be done in accordance with the main Neurolinguistic Programming principles.

The interconnections between the five systems and their elements are given in figure 2.

---

## Conclusion

The developed methodology of viable model-based enterprise management helps the modern enterprises to survive and growth in a highly dynamic environment. It uses the last achievements in the information technology at the current moment. To make this project complete we need to build a new type of Neural Networks which can work in a multitasking mode with dynamic weights generation, short and long term memory management properties and high adaptation and transformation capabilities. Such kind of networks is viable and can survive in an environment with a high degree of complexity and uncertainty because of their ability for self-development.

---

## Bibliography

- [Недева В., 2003] Недева,В., Един подход за изграждане на маркетингова информационна система от интегриран тип, Научни публикации "Техномат & Инфотел", Том 4, част1, 2003 г., Issue of Science Invest Ltd., Brunch Bourgas, Bulgaria, ISBN 954 909 72 8 5, стр. 314-326
- [Beer S., 1956] Beer S., Decision and Control. Great Britain, John Wiley and Sons Ltd, 1956
- [Beer S., 1959] Beer S., Cybernetics and Management. Oxford, English Universities Press, 1959
- [Beer S., 1967] Beer S., Cybernetics and Management. London, The English Universities Press LTD, 1967
- [Beer S., 1979] Beer S., The Heart of the Enterprise. New York, John Wiley and Sons, 1979
- [Beer S., 1981] Beer, S., Brain of the Firm. New York, John Wiley and Sons, 1981
- [Beer S., 1984] Beer, S., "The Viable System Model: Its Provenance, Development, Methodology and Pathology." Journal of the Operational Research Society 35(1): 7-25., 1984
- [Beer S., 1985] Beer, S., Diagnosing the System for Organizations. Great Britain, John Wiley and Sons Ltd, 1985
- [Kovacheva T., 2004] Extended Executive Information System, International Journal "Information Theories & Applications", Vol.11, Number 4, pp.394-400, 2004
- [Kovacheva T., Toshkova D., 2005] Kovacheva T., Toshkova D., Neural Network Based Approach For Developing The Enterprise Strategy, KDS-2005 (in print)

---

## Author's Information

**Todorka Kovacheva** – Economical University of Varna, Bulgaria, Kniaz Boris Str,  
e-mail: [todorka\\_kovacheva@yahoo.com](mailto:todorka_kovacheva@yahoo.com), phone: +359899920659

## ANALYSIS OF MOVEMENT OF FINANCIAL FLOWS OF ECONOMICAL AGENTS AS THE BASIS FOR DESIGNING THE SYSTEM OF ECONOMICAL SECURITY (GENERAL CONCEPTION)

**Alexander Kuzemin, Vyacheslav Liashenko, Elena Bulavina, Asanbek Torojev**

**Abstract:** *Some directions in the financial flows stable functioning are analyzed. The method of attack of the financial flows mutual action in different countries is offered. The main components of the financial flows' investigation, their stability and possible ruptures from the standpoint of the adequate economical security system design are substantiated.*

**Keywords:** *financial flows, economical security, fund market, fund indices, bank activity.*

---

### Introduction

---

Investigation of the problems associated with economical security of different subjects of management (starting with separate enterprises and finishing with regions, the country as a whole) has constantly been the focus of attention. This is associated with

- transient economical processes, in some instances this borders on unforeseen actions and errors in forecasts as to the subsequent economical development;
- prevailing of the globalization tendency in formation of various processes both of purely economical nature and different spheres of scientific and engineering development;
- a scale of emergencies action on all spheres of human activity and with the necessity to realize the assumed plans completely;
- possibility to consider unforeseen economical crises and recessions as specific situational aspects of management affecting the efficiency and effectiveness of the decisions being made.

In this case the circle of the mentioned above problems manifests itself the most acutely at a period of transformation changes which by their nature embrace permanently all institutional formations without exception both in the developed and developing countries due to the evolution process of the economical relations development [1]. Such a situation emerges because at the period of transformation changes the probability of ruptures between the flows of the real and financial sectors economy increases. That is why, in our opinion, the priority of definite investigations in the framework of the indicated direction is associated with revealing of regularities in movement of financial flows of different economical agents, this can be defined more concrete on the basis of the corresponding publications analysis.

---

### Analysis of Publications and Substantiation of the Investigation Objectives and Problems

---

First and foremost the analysis of the works devoted to problem under consideration points to wide spectrum of different opinions and directions of investigation.

So, in particular, A.Gorbunov considers the problems of managing financial and goods flows of economic organizations in the context of the imitation simulation [2]. Solving in this case mainly the problems associated with visualization and semantic interpretation of the investigated flows' movement offering tools of an illustrative representation of the flow processes. V.V. Kornejev analyzes interconnection of credit and investment flows of funds in the financial markets [3]. D.O. Chukhlantsev investigates interconnection of financial flows in the framework of corporative structures [4]. But despite the importance of the obtained results the general methods of the financial flows' analysis are missing from the works of the cited authors, this makes it difficult to formalize available results for implementation in the economical security system. In this case the analysis of the financial flows movement based on the integral matrices of financial flows incorporating uninterrupted movement of the financial resources of different economical agents from the moment of their incomes formation to the final usage

is the most formal one [5, 6]. Nevertheless, not all the questions associated with the financial flows movement fall within such matrices.

Primarily it is concerned with the analysis of the current tendencies directly associated with the economical security estimation. Thus, we shall separate as the main investigation problem the necessity to study a circle of questions making it possible to estimate the efficiency of the financial flows movement in the current prospects and substantiate the key aspects of their analysis from the standpoint of the efficient security system creation.

### **Analysis of Financial Flows Movement from the Outer Positions**

First of all, let us take into consideration that the majority of different managing subjects of economy experience the influence of the globalization processes taking place in the modern economical space, the degree of the action is the first level of the financial flows movement efficiency presentation.

To reveal one of the components of such dependence it is pertinent to analyze the dynamics of indices classifying the outer action on the financial flows movement efficiency. As such index, for example, we may chose EMBI+(The Emerging Markets Bond Index Plus) which is calculated in J.P. Morgan Chase Bank and which characterizes, in a way, the general investment appeal of a separate country (the smaller is the given indication, the greater attractive are securities of the given country, so the corresponding flows of financial resources seem more stable, the probability of rupture between the flows of the real and financial sectors of economy is predicted as less probable). Fig.1 shows the dynamics of the given index in section of separate countries (according to [www.cbonds.info](http://www.cbonds.info) data).

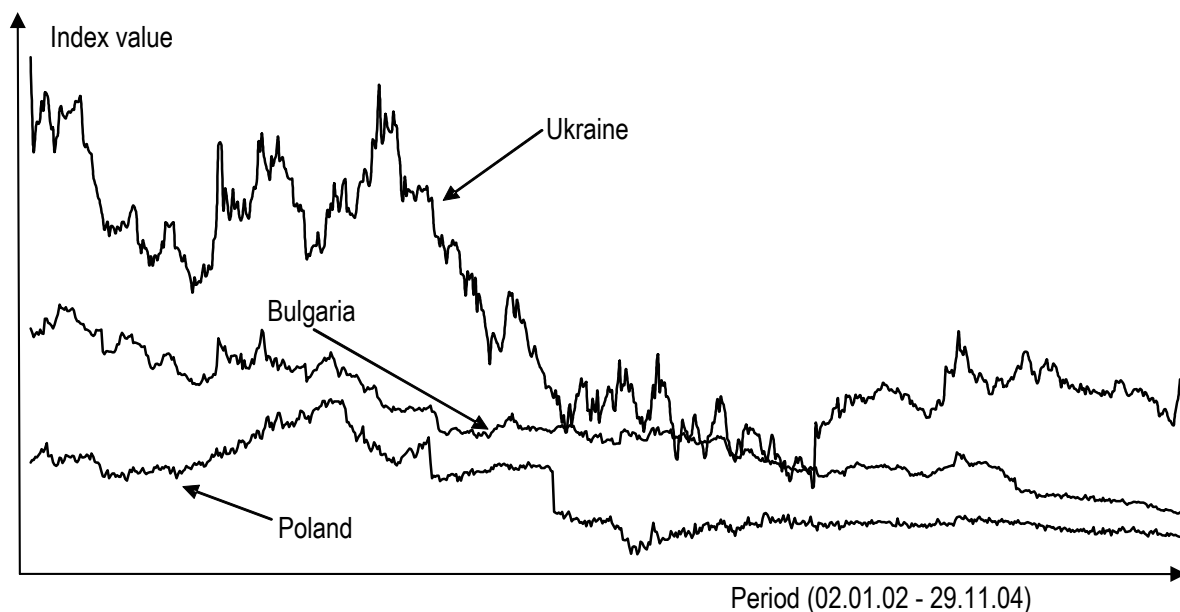


Fig.1. Dynamics of index EMBI+ for some countries at the period from 02.01.02 till 29.11.04

As evident from the data in Fig.1 in recent years one can observe a visual identity in the tendency of development of the index under investigation for such countries as Poland and Bulgaria, this allows to speak about possible similar tendencies in their financial flows movement. Nevertheless, to make more significant conclusions it is necessary to have quantitative estimates of such dependence. Table 1 lists the matrix of the correlation interconnection between the temporal lines of values of the index EMBI+ for some developing countries in the world. The analysis of the data from Table1 points to the availability of similar tendencies in the financial flows

movement for the majority of the countries being analyzed. Perhaps, Argentina is an exception by virtue of the greatest in history default experienced in recent times.

Alongside one can notice that for some countries the mutual correlation of the index is rather high, this points in our opinion to possible mutual influence of the financial flows of these countries on each other. As an example we refer to the dynamics of values of the index EMBI+ for Ukraine and Russia (Fig.2 according to the data from [www.cbonds.info](http://www.cbonds.info)) and Colombia and Mexico (Fig.3 according to the data from [www.cbonds.info](http://www.cbonds.info)).

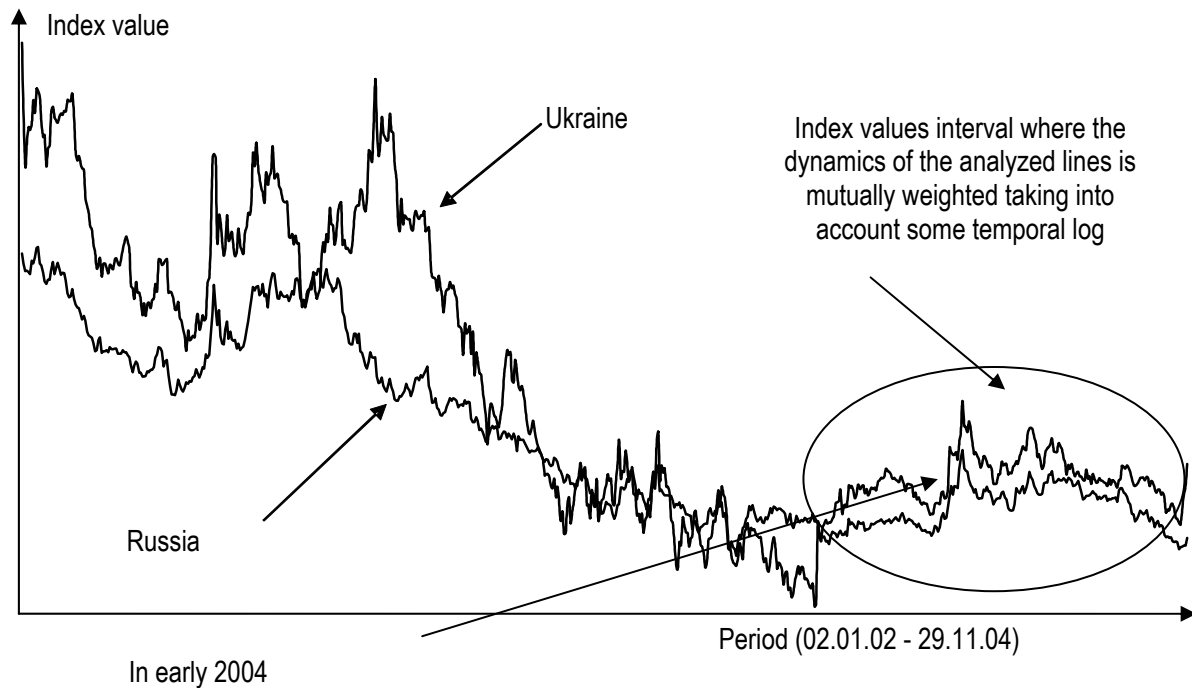


Fig.2 Comparison dynamics of the indices EMBI+ Ukraine and EMBI+Russia

Table 1.

Matrix of the correlation interrelation between temporal lines of values of index EMBI+ for some countries

Countries	Ukraine	Argentina	Brazil	Bulgaria	Venezuela	Columbia	Mexico	Poland	Russia	Turkey
Ukraine	1									
Argentina	0,217208	1								
Brazil	0,713859	0,663213	1							
Bulgaria	0,837146	0,110292	0,610115	1						
Venezuela	0,601131	0,380723	0,670548	0,6894	1					
Columbia	0,726216	0,574627	0,940122	0,656976	0,645966	1				
Mexico	0,794158	0,571706	0,950982	0,731471	0,774082	0,953299	1			
Poland	0,859167	0,508507	0,866693	0,806722	0,707537	0,907848	0,927919	1		
Russia	0,937923	0,253089	0,755033	0,906868	0,659145	0,807702	0,858808	0,91792	1	
Turkey	0,650442	0,449967	0,83776	0,731123	0,86146	0,812148	0,880083	0,803806	0,75216	1

As evident from the data in Fig.2 and Fig.3, there are definite intervals where the analyzed indices lines are identical respectively to some group of transformations of their temporal variations. Here, both in the first case and in the second case the index line being placed higher can be considered to be subordinate to the lower line i.e. the line which is in the given case the dominating one (based upon the efficiency of dynamics of index EMBI+).

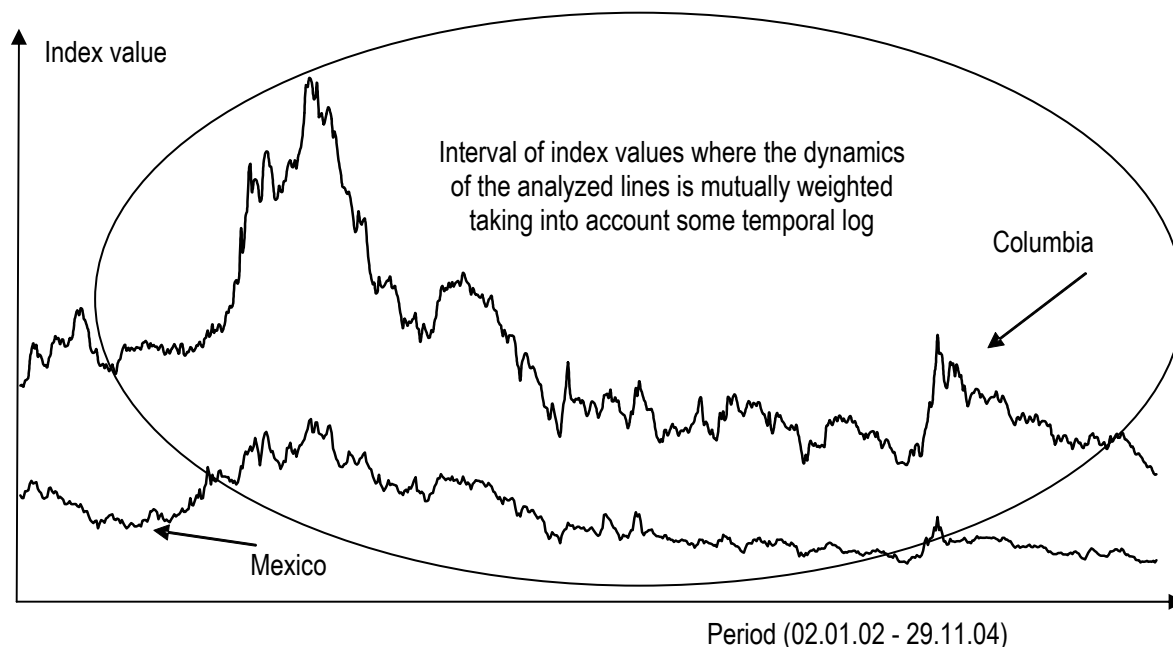


Fig.3. Comparative dynamics of the index EMBI+Columbia and EMBI+Mexico

From the above reasoning we can formulate the following hypothesis: if there is subordination between two lines manifestation of which is the stability of the correlation interrelation with regard to a definite log and a group of line transformations then it should be spoken about mutual influence of one line on the other one. Here, two cases are possible:

- a country with a dominating line exerts harmful effect upon the financial flows movement of the country of subordination line if there is a stable correlation interrelation in the positive logs of the lines;
- a country with a dominating line exerts positive effect upon the financial flows movement of the country of the subordination line if there is a stable correlation interrelation in the negative logs of the lines.

### Financial Flows Movement with Regard to the Internal Factors

The second direction of the financial flows dynamics investigation can be analysis of some indices of bank activity which incorporate the generalized effect of the outer factors. Among the variety of such indices in our opinion we should, primarily, set apart such as nonproductive assets in the volume of the granted credits and general dynamics of the internal credits gain (Table 2, according to the data from [www.ebr.com](http://www.ebr.com)).

The validity of separation of such indices in our opinion is associated with the fact that they characterize the most comprehensively the rupture of flows of the real and financial sectors of economy. In this way nonproductive credits directly influence the stability of the financial flows movement as these credits cover those categories of credits classified as non-qualitative, impaired and unrequited ones.

In this case a swift increase in the given credits can be considered as an instability growth in the financial flows movement as the probability of credit risks increases in the case of deterioration of macroeconomic situation in the country.

Table 2.  
Dynamics of some indices of financial flows movement in some NIS countries

Indices	Variation of indices year after year					
	1998	1999	2000	2001	2002	2003
Azerbaijan						
Nonproductive credits (% of the total sum of credits)	19,6	37,2	–	–	19,7	14,6
Dynamics of the internal credits (in % by the end of the year)	8,8	-10,4	13,5	-38,1	84,2	27,1
Kazakhstan						
Nonproductive credits (% of the total sum of credits)	–	-3,7	5,8	2,1	2,0	2,1
Dynamics of the internal credits (in % by the end of the year)	38,6	35,4	57,3	17,1	30,2	24,1
Kirghizia						
Nonproductive credits (% of the total sum of credits)	0,2	6,4	16,3	13,8	13,3	11,2
Dynamics of the internal credits (in % by the end of the year)	22,6	5,0	10,0	-8,1	18,9	10,9
Russia						
Nonproductive credits (% of the total sum of credits)	30,9	28,1	16,1	12,2	11,4	10,4
Dynamics of the internal credits (in % by the end of the year)	71,0	36,1	12,1	27,0	26,5	26,5
Ukraine						
Nonproductive credits (% of the total sum of credits)	34,6	34,2	32,5	–	–	32,7
Dynamics of the internal credits (in % by the end of the year)	58,0	30,5	23,1	18,7	28,9	39,6

When the data from Table 2 and Fig.2 are compared, then it may be concluded that EMBI indices' deterioration for Ukraine at the beginning of 2004 was a result of a sharp increase in the given credits and the existing highly stable part of nonproductive credits, which, on the whole, increased the risk of ruptures origin between the financial flows of different economy sectors.

No less important in terms of the analysis of financial flows movement is the investigation into dynamics of the fund indices which are representative of the processes of redistribution of free monetary and financial resources. But in the given aspect attention should be paid, first of all to the complexity of performing the corresponding analysis due to the uncertainty of fund indices dynamics in the developing countries. A manifestation of such uncertainty is a fundamental distinction between the indices statistical characteristics of the developing countries and the standard values; this allows using the unified procedures of their analysis [7]. Nevertheless, the aspiration for the markets integration can influence significantly the dynamics of the financial flows movement.

It would appear natural that the factors considered above are not exhaustive when analyzing the dynamics of financial flows. Other circumstances influencing the financial flows stability may be considered alongside with the possibility of the emergence of various ruptures in their movement. Because of impossibility to embrace all the aspects of the given problems in the framework of one paper we think the considered above to be the most significant when developing economical security in terms of various subjects of management. Moreover, the presented material allows in our opinion to separate the most typical and conceptual moments of the financial flows analysis.

### **Characteristic Aspects of the Financial Flows' Analysis**

Thus, among the main components of investigation of the financial flows movement revealing of their stability and possible ruptures in terms of development of the adequate system of economical security it is necessary to single out the following:



- 
- firstly, creation of a subsystem for analysis of the external factors influencing the financial flows' movement on the basis of the indices system characterizing the position in the foreign market. In particular, it is expedient to investigate EMBI+ index as one of such analysis directions;
  - secondly, creation of a subsystem analyzing the degree and direction of the action of other countries and organizations on the financial flow;
  - thirdly, creation of a subsystem analyzing the financial flows direction based on the indices of the bank sector of economy taking into account the study of the possibility of emerging crisis situations with the mutual movement of the real and financial sectors of economy, influence of the foreign banking capital;
  - fourthly, creation and development of methods for analysis the monetary and financial resources redistribution efficiency through functioning of the stock exchanges, investigation into mutual tendencies in variation of indices in different countries, separate segments of the market.

---

## Conclusions

---

Consideration of any problems associated with economical security is actual and needed. In this connection selection of a definite indices system the most completely and mutually characterizing the financial resources movement, the degree of influence of different financial flows and possibility of ruptures origin between the flows of the real and financial sectors of economy is essential. At the same time, application of the corresponding formal mathematical apparatus, which should be singled out as one of the priority directions for further researches, will serve to achieve the goal.

---

## References

---

1. On the problem of theory and practice of the transient period economy (Transactions of the Academic Council of IMEMO RAS, June 3-5, 1996, under the chairmanship of academician V.A. Martynov) –M.: RAS IMEMO, 1996, -104p. (In Russ.)
2. Gorbunov A. Financial flows control. M.: Globus, 2003.-224p. (In Russ.)
3. Kornejev V.V. Credit and investment flows of capital in financial markets. Monograph. –K.: NDFI, 2003. 376p. (In Ukr.)
4. Chkhlantsev D.O. Simulation of financial flows movements in the vertically integrated company and rationalization of its mutual settlements with contracting parties //Economic cybernetics: methods and means of effective management. – Perm: PSU, 2000. –P.213-219. (In Russ.)
5. Moudud J.K. Finance in a Classical and Harrodian Cyclical Growth Model. –The Jerome Levy Economics Institute.- Working Paper.- 2001. –52p.
6. Evstignejev V.R. Financial market in the transient economy. – M.: Editorial URSS, 200. –240p. (In Russ.)
7. KaminskyA., Shpirko O. Peculiarities of income distribution in new European stocks markets //Bankivska Sprava. – 2003.-№№5,6.

---

## Authors' Information

---

**Kuzemin A.Ya.** – Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, (Ukraine), [kuzy@kture.kharkov.ua](mailto:kuzy@kture.kharkov.ua)

**Liashenko V.V.** – senior scientific employee Kharkov National University of Radio Electronics, (Ukraine), [kuzy@kture.kharkov.ua](mailto:kuzy@kture.kharkov.ua)

**Bulavina E.S.** – senior scientific employee Kharkov National University of Radio Electronics, (Ukraine), [kuzy@kture.kharkov.ua](mailto:kuzy@kture.kharkov.ua)

**Torojev A.A.** – General Director of the Joint Russian-Kirghiz Venture “Computing Technique and Automation Means”

## USING ORG-MASTER FOR KNOWLEDGE BASED ORGANIZATIONAL CHANGE

**Dmitry Kudryavtsev, Lev Grigoriev, Valentina Kislova, Alexey Zablotsky**

**Abstract:** Enterprises in growing markets with transitional economy nowadays encounter extreme necessity to change their structures and improve business processes. In order to support knowledge processes within organizational change initiative enterprises can use business modeling tools. On one hand software vendors suggest many tools of this kind, but on the other hand growing markets with transitional economy determine quite special requirements for such tools. This article reveals these requirements, assesses existing business modeling tools using these requirements and describes ORG-Master as a tool specially created for support of process improvement initiatives in the growing markets with transitional economy.

**Keywords:** Business information modeling, business modeling, knowledge process, organizational change, business process improvement, growing markets, transitional economy.

---

### Introduction

---

ORG-Master is a business modeling software, which was initially created as a response to growing need for computer aid to consulting projects in the field of organizational change and business transformation. In spite of the diversity of products for business modeling ORG-Master has certain advantages that can be revealed in solving certain tasks in certain environment.

Certain tasks include such organizational change components as business process improvement, business restructuring, quality management implementation and holistic improvement of management system. In the current article, organizational development will be described by the example of business process improvement (BPI) initiative.

Certain environment includes growing markets with transitional economy (GMwTE) which determine specialties in organizational change initiatives. GMwTE include post-soviet countries (Russia, Ukraine, Belarus, Kazakhstan) and in the current article will be described by the example of Russia. In order to reveal these specialties Section 1 describes features of GMwTE from management point of view. Section 2 focuses on the flow of knowledge within BPI initiative and gives an ability to define requirements for business modeling tool at the GMwTE (section 3). Section 4 reveal imperfections of existing business modeling tools with respect to above-mentioned requirements and show the niche for ORG-Master. Section 5 explains the main concepts and consequent advantages of ORG-Master. Section 6 describes practical application of ORG-Master.

---

### 1. Business process improvement initiatives in the growing markets with transitional economy.

---

The most important features of GMwTE from management point of view are:

1. Extremely high pace of change in market conditions and business environment
2. Low level of managerial culture
3. Predominance of informal methods of management

Quick changes and competition growth make companies to change in the same pace and the main objectives in the organizational change is to fit company structure with business needs and to implement client-oriented business processes that allow to achieve company goals. This results in the necessity to launch restructuring or BPI initiatives.

The main prerequisite for BPI initiative is transparent management at every level of organization. In this context transparency implies holistic knowledge describing *What* functions and processes are realized in the company, *Who* performs the functions, *How* the functions are performed, *What for* are the functions performed. While low level of managerial culture results in absence of clear knowledge in this field. As a results BPI initiative in the

GMwTE usually involve a wide range of preliminary stages directed towards understanding of company “big picture” in order to make conceptual changes and define the processes for improvement or re-engineering.

The third feature of GMwTE - predominance of informal methods of management results in small amount of documents and business rules. Such a situation has its roots either in skeptical attitude to archaic and out-of-date formal documents at post-soviet enterprises or in quick growth of small start-ups. In some situations, informal intuitive method of management brings fruits, but it is terminated by the scale of business and is one of the barriers in development of managerial culture. As a result, BPI initiative in the GMwTE has an important objective – to switch company from informal methods of management to formal procedures and business rules.

## 2. Knowledge process in the business process improvement initiative.

BPI or restructuring initiatives deal with business organization knowledge. Under *business organization knowledge* in the current article we will understand *knowledge domain* covering organizational goals, structure, processes, functions, rules, rights, authorities and relationships between this objects. Thus in order to raise effectiveness of BPI initiative project team should support knowledge process in the domain of business organization knowledge. As described in [Strohmaier, 03a] knowledge infrastructure<sup>1</sup> is determined by the nature of knowledge process, which in turn can be understood through analysis of business processes covered by improvement initiative (figure 1).

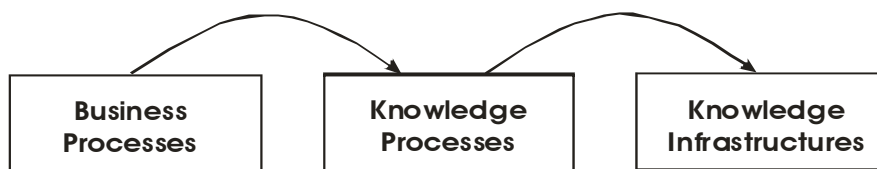


Figure 1: Business process and knowledge infrastructure relationship

The main constituents of improvement initiatives are *organizational change processes* - a subset of the whole system of business processes:

- business process analysis
- business process improvement
- organizational structure control
- performance management

Business analysts (either internal or external consultants) together with domain experts (head of departments and other managers) *generate* business organization knowledge, *store* and *transfer* it throughout the organization in these processes.

Application of business organization knowledge is distributed between all the other business processes - operating, management and support processes. Organizational roles of performers vary from workers to executives (top managers).

According to [Strohmaier, 03b] business organization knowledge can be visualized (figure 2).

The most important and influential feature of this process consists in different organizational roles involved in it and especially in the knowledge transfer process. During transfer process business analysts deliver their knowledge through the mediation of domain experts to personnel from different domains and organizational levels. As it was mentioned in [Section 1], one of the goals of BPI initiative in the GMwTE is to shift the priorities of management from informal methods to formal business rules. Thus the basis of knowledge transfer

<sup>1</sup> Under Knowledge Infrastructure we imply all the means that enable effective knowledge management within organization ~ knowledge process support

is formalized knowledge and the main factor of its successful internalization [see Nonaka, 03] by personnel is type of knowledge representation.

Type of knowledge representation depends on two specific knowledge processes generation on one hand and application on the other. While the way these processes are performed is determined by the involved organizational roles.

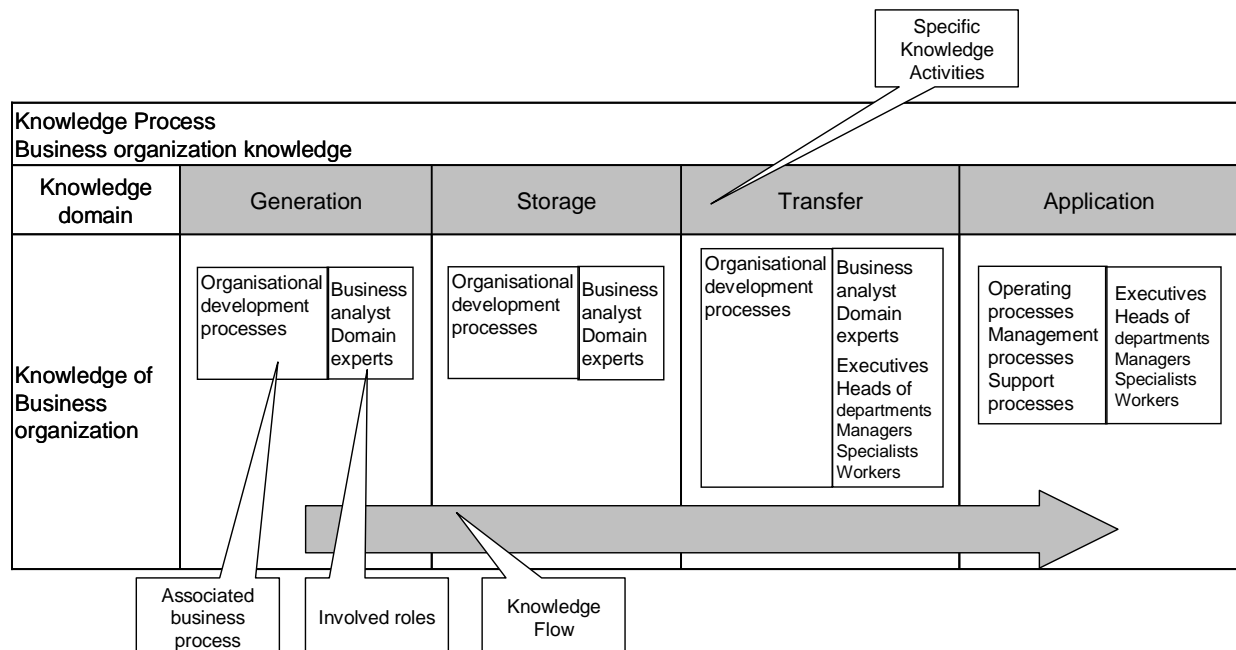


Figure 2: Knowledge process in the business improvement initiative

Business analysts have developed competencies in organizational management and system analysis. They have detailed understanding of business from different points of view and can operate with different objects and their relations (organizational units, functions, processes, goals etc).

Personnel from different domains and organizational levels have low competencies in organizational management see [Section 1]. They usually have only dim understanding of business from "organizational unit" point of view (answer for the question "who do what?").

As a result, business analysts primary use diagrams of different types and notations (IDEF0, UML, EPC) as a mean of knowledge representation. The other personnel use job descriptions, documents describing the functions of business units / departments and other regulating documents. In addition, these regulating documents for application usually prepared according to national, industrial or corporate standard. Namely these regulating documents solve one of the objectives of BPI initiative in the GMwTE - shift the priorities of management from informal methods to formal business rules see [Section 1].

Thus, the necessity to facilitate communication between people speaking "different languages" predetermines important requirement for knowledge infrastructure.

### 3. Requirements to business modeling tool for GMwTE.

For the current analysis we assume business modeling tool primary as a support system for knowledge generation and storage during BPI initiative.

Previous section described the necessity to have different types of knowledge representation during BPI initiative in the GMwTE. Assumption that the process of knowledge transfer do not change the type of knowledge representation imply the necessity to generate knowledge both in type of diagrams for analysis and regulating

documents for application. This requirement for knowledge generation process determine the first requirement for business modeling tool:

**1. Ability to represent knowledge in different types and formats**

Section 1 highlights the necessity of preliminary stages within BPI initiative in the GMwTE. For example, companies should define goals, composition of functions, change organizational structure, reassign responsibilities for function realization, reveal a list of business processes. This tasks can be done both in series and in parallel and include several analysts concentrating either on different tasks or on different levels of detail. Such a nature of BPI initiative determine the next requirement:

**2. Ability to work both with a complex model (e.g. business process model) and with separate parts of this model (relate functions with organization roles, roles with infrastructure etc) using different views of enterprise.**

Fast dynamic of the enterprise development is especially relevant for GMwTE and require constant improvements in business processes thus a model once created should be constantly up-dated. Model is a system of constituent objects and their relationships, but both objects and their relations change constantly. This situation generate the third requirement:

**3. Ability to reflect changes in objects and in their relationships throughout the whole model after changing any part of the model.**

In order to reveal a tool, which satisfy all the requirements mentioned above an analysis of the tools existent in the Russian market was carried out.

---

#### **4. Analysis of existing business modeling tools in the Russian market**

---

Although in some BPI initiatives knowledge is created and stored using typical office applications like MS Word or Excel or simple graphical packages like MS Visio this tools obviously do not satisfy requirements see [Section 3].

The main business modeling tools existent in the Russian market that are usually used for organizational development and BPI purposes are:

ARIS <http://www.ids-scheer.com/>

BPWin (AllFusion Modeling Suite) <http://ca.com/>

There are also some Russian products that contain either limited functionality or slight modifications of foregoing tools. Differences of these products are immaterial from point of view of chosen requirements and as a result, they appeared beyond the scope of our analysis.

There are also a broad range of CASE tools (e.g. Rational Rose) for corporate systems development. These tools include business process modeling, but their primary function is information architecture development and it determines their whole viewpoint for enterprise modeling. As a result they are nor convenient for organizational management and business process modeling, nor efficient. Thus, they appeared beyond the scope of our analysis.

Here is generalized result of the analysis:

**Requirement 1:** Ability to represent knowledge in different types and formats

ARIS: It includes a broad library of object types and corresponding diagrams, but it has a very complicated mechanism for generating regulating documents. It is hard to customize necessary templates and consequently requires unique and expensive specialists

BPWin: It allows generating IDEF0 diagrams, but it is also very hard to generate corresponding regulating documents in customary standards.

**Requirement 2:** Ability to work both with a complex model (e.g. business process model) and with separate parts of this model

ARIS: Satisfy. There are both a whole process model and separate constituent models.

BPWin: Dissatisfy. User works either with one object type (functions, roles) or with a whole model of business process (one type of composite diagram).

**Requirement 3:** Ability to reflect changes in objects and in their relationships throughout the whole model after changing any part of the model.

ARIS: Partially. Centralized library of modeling objects guarantee the reflection of changes in the particular object throughout the model (e.g. changing function name in one diagram cause changing this name in every diagram in the model), but changes in relationships between objects of different type do not appear automatically throughout all diagrams.

BPWin: Satisfy. All the objects stored in centralized library and are used in one type of diagram.

Thus, presented tools do not completely satisfy suggested requirements. Besides this tools are quite expensive and require extremely professional analysts to support business model.

There is a necessity for more effective business modeling tool for organizational development.

## 5. Main concepts and advantages of ORG-Master

### Concepts and methodology

The main idea of ORG-Master consists in division of business modeling interface from model representation one. As a result, each interface and type of knowledge representation is optimized for the solution of own tasks. This idea is contrary to an approach of ARIS and BPWin. In the foregoing product user input, editing and represent business model in the same knowledge representation type and format.

Division of interfaces in ORG-Master allows representing knowledge both in different types (diagrams in different notations, reports, tables) and from different point of views.

On the other hand business model editing interface has its own type of knowledge representation based on two instruments: classifier (ontological models, see [Gavrilova, 00]) and matrix (table).

Classifier – hierarchical tree of particular objects (e.g. organizational roles, functions, material resources, documents etc), that can have different attributes: type, meaning, comments etc. In the process of building classifier objects become structured into a hierarchy/ tree – they receives relationships of AKO (“A Kind Of” [Gavrilova, 00]) type (figure 3).

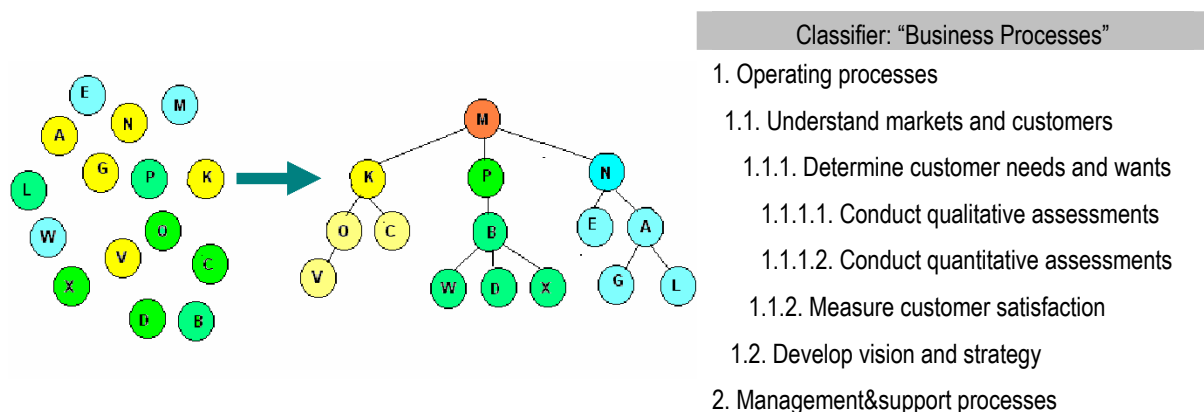


Figure 3 Structuring informational objects

Matrix (table) – models that define relationships between objects of different classifiers in any combination of the later (figure 4). Relationships can also have different attributes (directions, type, name, index, meaning).

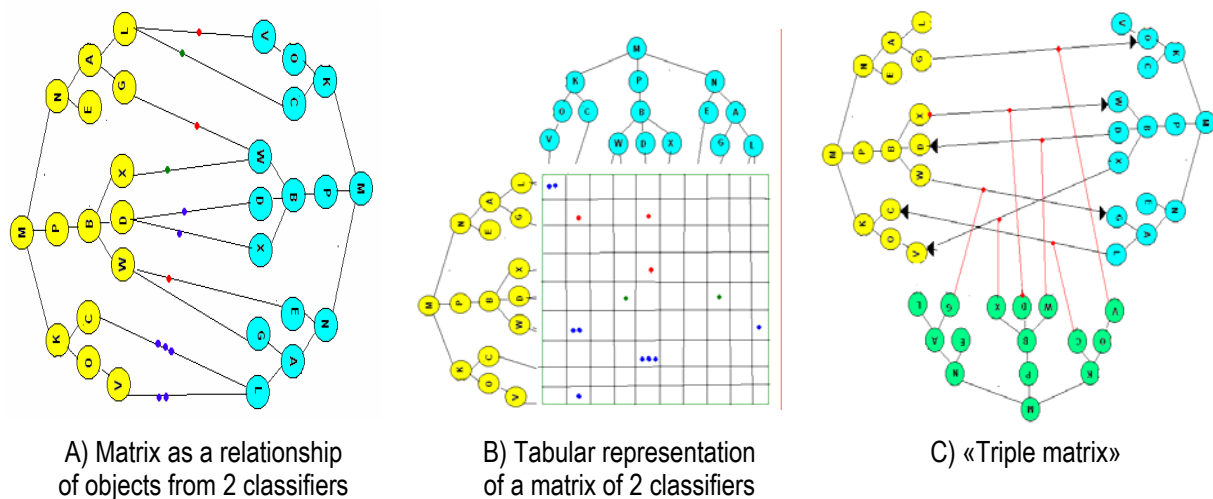


Figure 4: Conceptual framework of matrix (table)

As any material object of any complexity (e.g. building) can be described using definite number of 2-dimensional (flat) schemes (e.g. design drawings) so and several matrix allow to receive multidimensional description of complex business system and make it both holistic and visible (see figure 5).

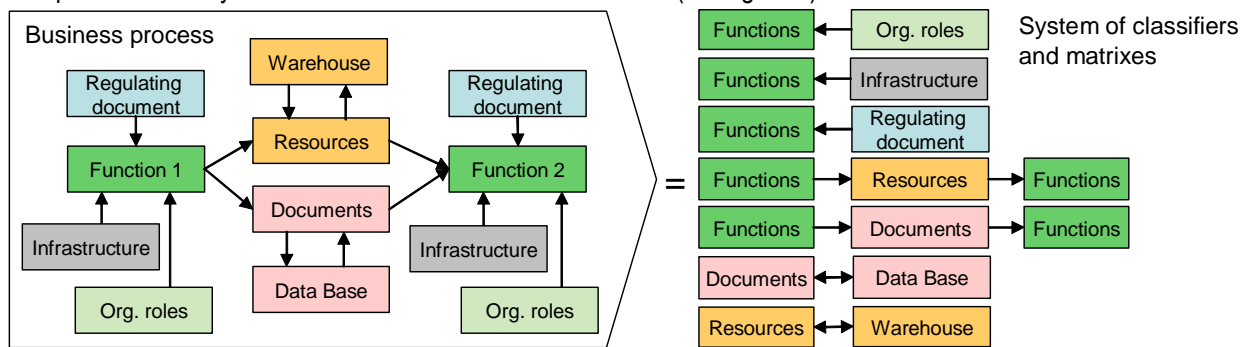


Figure 5: Business process model as a system of classifiers and matrixes

**ORG-Master advantages for end user**

Foregoing concepts of ORG-Master provide the following features of this tool:

- ability to generate multidimensional reports based on matrixes from business model with different level of detail, which allow to analyze company from many viewpoints for people at different levels of organizational hierarchy
- ability to fine-tune knowledge representation reports, that allow to generate regulating reporting on the basis of business model for particular needs and customary standards
- ability to generate visual diagrams of business processes that support business analysis
- ability to decompose the whole business model into constituent separate submodels, which allow to divide complex BPI initiative into manageable tasks and solve problems in separate domains with adequate (pared-down) tool
- all the objects and relationships from different submodels are integrated into centralized holistic model that allow to reflect changes in objects and in their relationships throughout the whole model after changing any part of the model

These features of ORG-Master satisfy requirements to business modeling tool for GMwTE and together with relatively low price and low training complexity characterize it as effective and efficient tool.

Among the relative disadvantages of this tool the most obvious is absence of quantitative analysis of business processes, but this feature is of low importance with respect to foregoing requirements.

---

## 6. Application of ORG-Master and practical results

---

ORG-Master has 6-year history of application in the organizational change and BPI initiatives. There is a broad range of ORG-Master clients located in Russia, Ukraine. Size of ORG-Master clients vary from small companies to large holding structures (up to 10000 people).

The results of typical ORG-Master application in BPI initiatives include:

- Business model which describes functions, organizational roles, goals and measures, function distribution among organizational roles and description of necessary business processes.
- Regulating documents based on business model (job descriptions, procedures etc)
- Diagrams of the necessary processes based on business model

---

## Conclusions

---

Since organizational change and BPI initiatives become a life-style of every company, it is useful to support such an activity with adequate tools for business modeling. However, choice of the tool is determined by the objectives of tool application and business environment. Current paper revealed specialties of BPI initiatives in the GMwTE and analyzed existing business modeling tools from that perspective. Because of this analysis, ORG-Master can be considered as an effective and efficient for knowledge process support during organizational change initiatives in the GMwTE.

---

## Acknowledgements

---

The author of this paper is thankful to the advisor Dr. Prof. Tatiana Gavrilova (St. Petersburg State Polytechnic University) for her useful suggestions about content and destiny of this paper.

---

## Bibliography

---

- [Strohmaier, 03a] M. Strohmaier Designing Business Process Oriented Knowledge Infrastructures Proceedings der GI Workshopwoche, Workshop der Fachgruppe Wissensmanagement, Karlsruhe (2003)
- [Strohmaier, 03b] M. Strohmaier A Business Process oriented Approach for the Identification and Support of organizational Knowledge Processes Proceedings der 4. Oldenburger Fachtagung Wissensmanagement, Oldenburg (2003)
- [BIG, 96] BIG&Expert "Seven notes of management", Moscow (1996)
- [Nonaka, 95] Nonaka I., Takeuchi H.: "The knowledge creating company"; Oxford University Press (1995)
- [Bukowitz, 99] Bukowitz W., Williams R.: "The knowledge management fieldbook"; Prentice hall, Pearson Education Limited (1999)
- [APQC, 96] APQC's International Benchmarking Clearinghouse Process Classification Framework [www.apqc.org](http://www.apqc.org), (1996)
- [Gorelik, 01] Gorelik S., "Business-engineering and management of organizational change"; (2001) <http://www.big.spb.ru/publications/bigspb/metodology/>
- [Gavrilova, 00] Gavrilova T., Horoshevsky V. "Knowledge bases of intellectual systems"; Piter / Saint-Petersburg (2000)
- [Рубцов, 99] Рубцов С., Сравнительный анализ и выбор средств инструментальной поддержки организационного проектирования и реинжиниринга бизнес процессов <http://or-rsv.narod.ru/Articles/Aris-IDEF.htm>
- [Репин, 01] Репин В. Сравнительный анализ нотаций. <http://www.interface.ru/fset.asp?Url=/ca/an/danaris1.htm>

---

## Authors' Information

---

**Dmitry Kudryavtsev** – Saint-Petersburg State Polytechnical University, Tkachey str., 24-24, Saint-Petersburg - 193029, Russia; e-mail: [dk@big.spb.ru](mailto:dk@big.spb.ru)

**Lev Grigoriev** – BIG-Petersburg (consulting company), Sovetskaya str., 2, Saint-Petersburg - 191014, Russia; e-mail: [spbbig@infopro.spb.su](mailto:spbbig@infopro.spb.su)

**Valentina Kislova** – BIG-Petersburg (consulting company), Sovetskaya str., 2, Saint-Petersburg - 191014, Russia; e-mail: [valya@big.spb.ru](mailto:valya@big.spb.ru)

**Alexey Zablotsky** – BIG-Petersburg (consulting company), Sovetskaya str., 2, Saint-Petersburg - 191014, Russia; e-mail: [support@big.spb.ru](mailto:support@big.spb.ru)



---

---

## INFRAWEB Project

---

---

### SEMANTIC WEB SERVICE DEVELOPMENT ON THE BASE OF KNOWLEDGE MANAGEMENT LAYER - INFRAWEB APPROACH

Joachim Nern, Tatiana Atanasova, Gennady Agre, András Micsik,  
László Kovács, Janne Saarela, Timo Westkaemper

**Abstract:** *The paper gives an overview about the ongoing FP6-IST INFRAWEB project and describes the main layers and software components embedded in an application oriented realisation framework. An important part of INFRAWEB is a Semantic Web Unit (SWU) – a collaboration platform and interoperable middleware for ontology-based handling and maintaining of SWS. The framework provides knowledge about a specific domain and relies on ontologies to structure and exchange this knowledge to semantic service development modules. INFRAWEB Designer and Composer are sub-modules of SWU responsible for creating Semantic Web Services using Case-Based Reasoning approach. The service and user agent (SUA) unit is responsible for building up the communication channels between users and various other modules. It serves as a generic middleware for deployment of Semantic Web Services. This software toolset provides a development framework for creating and maintaining the full-life-cycle of Semantic Web Services with specific application support.*

**Keywords:** *Semantic Web Services, Fuzzy Set, Ontologies, Case-Based Reasoning, Multi-Agent Systems*

---

#### Introduction

---

The primary objective of the INFRAWEB project is to develop an ICT framework consisting of several specific software tools, which enables software and service providers to generate and establish open and extensible development platforms for Semantic Web Service based applications [Nern, 2004]. This software tool set facilitates the establishment of virtual development platforms as well as interoperable middleware designed for a semantic and ontology-based handling of Semantic Web Services oriented on given conception WSMO specifications.

Generated in this way, the open platform (Fig. 1) consists of coupled and linked INFRAWEB units, whereby each unit provides tools and system components to analyse, design and maintain WEB-Services realised as Semantic-Web-Services within the whole life cycle.

As illustrated in Fig.1 the overall design is structured in three main layers:

- 1) a knowledge management layer for handling service related knowledge artefacts realised as an organisational memory coupled to semantic information routing components (OM&SIR),
- 2) a service development layer for creating and maintaining Semantic Web Services embedded in a semantic based interoperable middleware, consisting of Semantic Web Service Designer & Composer, Distributed Semantic Web Service Registries, and an agent based discovery module (Semantic Web Service Unit -SWU)
- 3) a service deployment layer for the execution and monitoring of Semantic Web Services exploiting closed loop feedback information (Quality of Service brokering) provided for distributed decision support issues.

The software tool-set building components map the specific modules of the INFRAWEB framework. In the rest of the paper the knowledge management and service development layers are considered in detail.

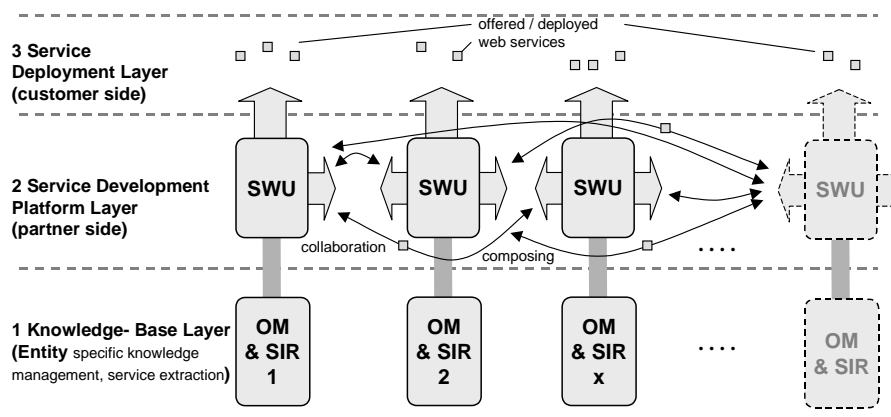


Fig. 1: Open and extensible development platform for the design, deployment and maintenance of Semantic Web Services as a net of coupled INFRAWEB units.

## Knowledge Management Layer

**Fuzzy Concept Based Organizational Memory - OM** As a repository for semi-structured knowledge artefacts a Fuzzy Set based organisational memory (FCM-OM) will be realised for knowledge management issues in the knowledge management layer [Nern, 2005a]. The OM management module is endowed with tools for knowledge acquisition and knowledge representation. This module is responsible for the collection, organisation, refinement, and distribution of knowledge objects handled and managed by the service providers. The current specification and realisation of the INFRAWEBs OM is based on the novel results of research activities in the area of Fuzzy Set theory:

Considering the ambiguity, imprecision of concepts (electronic knowledge objects of an entity, including the objects handled and received in Internet related environments) an useful approach is the adaptation and application of FCM (Fuzzy Conception Matching) methods [Zadeh, 2002]. Within this approach a “concept” is defined (and represented) by a sequence (a set) of weighted keywords [Zadeh, 2002]. Ambiguity in these concepts is defined by a set of imprecise concepts, whereas each imprecise concept is defined as a set of fuzzy concepts (using methods of implicit semantics), which is related to “a set of imprecise terms representing the context” [Zadeh, 2002].

Involving and considering also formal semantics, this imprecise terms (words) are “translated” into precise terms (words) formalised as an ontology [Zadeh, 2002].

Within the INFRAWEBs project two streams are focused in realising this system component:

- one component of the OM (FCM-OM) is designed following the rules of implicit & soft semantics (using statistical based AI methods like clustering and classification)
- a second component (O-OM) reflects the Knowledge handling based on methods related to formal semantics (Ontologies) – hard semantics [Zadeh, 2002].

**Semantic Information Routing - SIR** A further module within the first platform layer is the Semantic Information Router (SIR) [Westkaemper, 2005] which is coupled to the OM.

The interface between the Semantic Information Router (SIR) and the OM is based on the SOAP protocol. The interface is used to provide WSDL based service descriptions to the OM component and to query for content item references that are related to given service description references.

The query interface of the SIR component is based on the SPARQL which is a RDF metadata query language designed by the W3C DAWG working group with protocol bindings defined for JDBC, HTTP and other protocol stacks. Technically the SPARQL interface is a metadata query language with syntax close to SQL which offers querying metadata as table and graph result sets. The protocol stack most suitable for J2EE environments is a JDBC type 4 (Java Database Connectivity) compliant driver. The implementation of a JDBC based protocol binding for SPARQL is named SPARQL4j (<http://sourceforge.net/projects/sparql4j>).

---

Registration of non-semantic atomic services is accomplished via a web-based GUI interface. The interface is mainly used to input WSDL and BPEL4WS based service descriptions and enter additional related non-functional properties. Additionally a UDDI registry interface is provided for the SIR component. This will be established via SOAP as a UDDI Subscription Listener.

---

### **Service Development Layer - SWU**

---

The main module within the second – the service development – layer is the Semantic Web Unit (SWU) [Atanasova, 2005], [Nern, 2005b], [Agre, 2005]. SWU provides knowledge about a specific domain and relies on ontologies to structure and exchange this knowledge.

The following challenges for developing of SWU have to be taken into account:

- Converting Web Services from available descriptions and domain knowledge (organizational memory) to the semantic ones;
- Composition of Web Services, combining and orchestrating them in order to deliver added-value services;
- Dependencies that arise when a service integrates with external services and becomes dependent on them;
- Integration Processes of several business partners situated on different locations that have to be integrated with each other.

SWS Unit ensures designing SWS from the domain knowledge. All knowledge objects from the organizational phenomena influent on the constructed semantic web service and consist of WSDL, BPEL4WS and UDDI files as web service descriptions together with WSML and metadata as ontologies and non-functional properties carriers.

One promising solution to SWS design is to define a library of reusable aspects that would allow the service developer to dynamically instantiate and configure all the needed aspects to deal with different SWS parts. These reusable aspects can be seen as generic templates that can be customized and integrated “on demand” to accommodate to service requirements. This consideration leads naturally to using Case-Based Reasoning approach for service development.

Within the SWU the Designer and Composer modules are responsible for decision supported creation of Semantic Web Services using the Case-Based Reasoning approach. The Designer is a tool for semiautomatic conversion of non-semantic Web services to Semantic Web Services, whereas the Composer enables the semiautomatic creation of new Semantic Web Services via the composition of existing Semantic Web Services.

The architecture of both modules is based on such general principles as:

- Specialization: Each tool is carefully designed based on analysis of specificity of the task it is intended to be used for. It leads to minimization of efforts the service provider should apply for creating a semantic web service. Such minimization is achieved via fully utilization of all available information resources about the service as well as CBR-based mechanism for improving the behaviour of a tool through accumulating and using experience of the service provider to work with this tool.
- User-friendliness: it is assumed that the users of our tools will be semantic Web service providers as well as customers of such services. In both cases the users will not be specialists in first-order logic that is why we implement a self-explained graphical way for constricting and editing of all elements of a semantic web service.
- Intensive use of ontologies: ontologies are the core concept of the Semantic Web technology; however, we consider that creating ontologies for different application domains requires very intensive cooperation of highly qualified domain knowledge engineers and logicians. Both categories of the users do not belong to the range of potential customers of our tool. That is why we assume that our customer will be mainly a user of already created ontologies rather than a creator of new ontologies. However, we foresee that in some cases the service providers have to be able to create some specialized versions of (general) existing ontologies. Means for creating such (restricted) ontologies are also included in our tools.
- Semantic consistency: operation with each tool is organized via ontology-based system-driven interaction with the service creator, which prevents him/her from possible errors and allows being concentrated on the relevant part of knowledge to be acquired. Application of context-sensitive syntactical and completeness

checks at each step of the semantic service creation prevents the user from constructing semantically inconsistent and incomplete models.

**Designer** With the Case-based Designer SWS a service provider creates semantic descriptions of the services on the base of set of ontologies, preferences (QoS) and business logic of services using service design templates (DST).

The Designer consists of several sub-modules responsible for WSMO compliant creation of main elements of INFRAWEBs specific Semantic Web Service.

SWS-Designer has to add the semantic meaning to Web Services about: Data, Functioning, Execution, Discovery, and Selection. This can be done by the following modules: Capabilities editor, BPEL-based editor, Grounding editor via using of DST with appropriate validation and indexing. Creating, storing, and retrieving of similar DST are organized using Case-based Reasoning (CBR) approach. A retrieved template can be further used or adapted by the user for designing the desired functional model of a new semantic Web service and/or to be stored in case-based memory for later re-use.

As a graphical user-friendly tool the Capability Editor facilitates construction and editing of complex WSML-based logical expressions used for representing service capabilities. The BPEL4WS-based editor serves for creating WSMO-based service choreography and orchestration. The Grounding Editor provides facilities for semiautomatic creating of WSMO-based grounding. A basic feature of the Designer is the use of Design Service Templates (DST), representing graphical models of capability and functionality (or their parts) of Semantic Web Services, which have been designed by the user in the past.

**Composer** With the Case-Based Semantic Web Service Composer a service provider constructs SWS semi-automatically in design-time by composing descriptions of existing SWS and using domain knowledge.

The SWS Composer uses the previous service compositions that form the general tasks. Such compositions are represented by Service Composition Templates (SCT).

The SWS Composer provides:

- Similarity-based retrieval of an appropriate semantic service template based on the description of capability of the desired service and description of its functionality
- Semi-automatic adaptation of service functional model based on the results of discovery of sub-services matching the template proxies
- Advertising the created service and its generalization and storing as new template for later re-use.

The Case Base of SCT consists of references to complex service scenarios constructed by SWS Composer in the past and associations between problem solutions (particular description of request for servicing made in the past) and founded solutions.

The Composer presents an interactive approach for composition of WSMO compliant Semantic Web Services. The SCT represents graphical models of the service composition as a control and a data flow between several semantic sub-services given by incomplete description of their capabilities. On the (Service) provider side the Composer enables the creation of a new composed "static" Semantic Web Service by discovering appropriate Semantic Services matching the required capabilities. Selecting of such services is implemented as an interactive system-driven semiautomatic process.

The Designer and Composer are implemented in the J2SDK 1.4.2 Runtime environment. Eclipse RCP (Rich Client Platform) is used for developing the basic platform components; plug-in infrastructure and graphical user interface components, whereas Eclipse GEF (Graphical Environment Framework) is the basis for implementing the graphical editors. Access to WSMO-based repositories (ontologies, Semantic Web Services, etc.) is realized via the WSMO API.

**WSMO API - WSMO4J** For ensuring compatibility and interoperability within and between the INFRAWEBs framework modules the WSMO API (WSMO4J) is applied. The WSMO4J is an open-source project (distributed under a LGPL licence) with two parts: a) WSMO API - application programming interfaces for WSMO, which allow basic manipulation of WSMO descriptions, e.g. creation, exploration, storage, retrieval, parsing, and serialization and b) WSMO4J - a reference implementation of the WSMO API, including a WSML parser.

One of the major advantages of using the WSMO API in INFRAWEBs is to assure the compatibility and interoperability between the SWS Designer and Composer modules, and the repository component. For example, the distributed SWS registry uses WSMO4J in the process of transforming WSMO element descriptions into RDF triples stored into an RDF triple repository for efficient query and management. Using WSMO4J enables easy integration and interoperability within the framework as well as with the WSMO Studio, thus some of the components can be realized as extensions (plug-ins) for the WSMO Studio.

**Agent Based SWS Discovery - SUA** The SUA (service and user agent) [Kovacs, 2005] unit is a basic layer of the INFRAWEBs software environment (Fig. 2). The Communicator is responsible for building up the communication channels between users and various other modules. The User circle denotes the application acting on behalf of the user: for example a GUI interface or an intelligent agent. The User circle and the SWS Composer need to discover and select existing web services with specified capabilities. The Discovery Agent supports this task, while the Service Agent supports the execution of the selected web service in cooperation with the SWS Executor module.

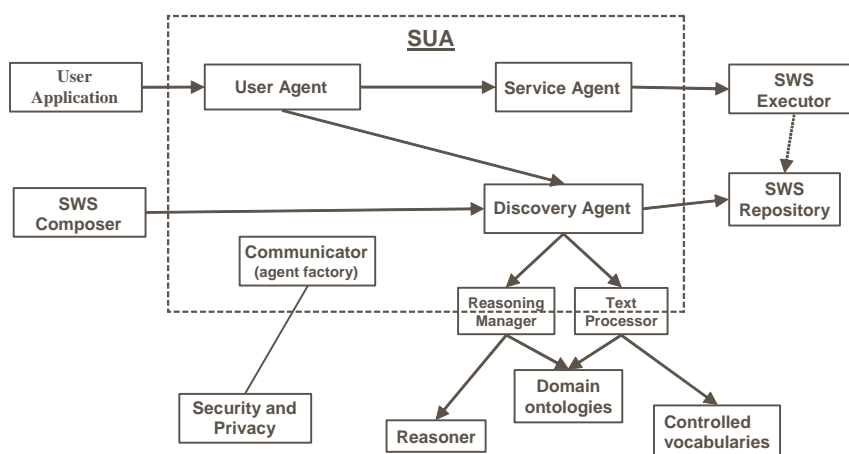


Fig. 2: Internal Architecture of Middleware Layer and its Dependencies with Other Modules

The discovery process is planned as a hybrid approach combining text processing and reasoning. Therefore, this process needs other external information and tools as well: domain ontologies contain the background information about service domains and organizational memories gathered from the INFRAWEBs framework and transformed into Semantic Web compatible format.

The SUA unit within INFRAWEBs acts as a middleware layer for user applications, connecting these applications to the functionality available in the form of Semantic Web Services. An important decision is whether to hide the semantic approach and accompanying logic-based framework of SWS from the user applications or not. An present investigation is to clarify: What level to choose for communication? On the level of plain web services, simple XML data structures are exchanged, which are easy to generate and process, but lack the possibilities of the semantic approach. On the level of semantic web services, facts and rules are exchanged in the form of logical expressions. As the latter case puts extra requirements for the user application, it is decided to find a middle course between the two solutions. Communication is kept on the semantic level, but its form is either hidden or aided with special functionalities and automatic translations.

Apart to SWS discovery the simplest way is based on keyword matching within the descriptive metadata of web services (similar to the UDDI approach). The most complex way is to logically prove that the web service is able to fulfil the given goal. Currently, response time of discovery process is a significant trade-off for these approaches.

The keyword-based method is improved if the goals and capabilities are used for keyword generation instead of metadata. Goals and capabilities are described using ontology terms, so more homogeneous and precise keywords can be extracted this way. The logic-based methods differ in aspects of goals and capabilities considered. It is simpler and faster to match only by post-conditions, but then matching services might not be executable because of missing or unacceptable information (input).

The project approach is to split the discovery into two phases. In the first phase, the keywords of capabilities are used to filter the possible web services. Then, logical matching is applied only for the filtered services.

The list of matching web services is returned to the user application enriched with descriptive and qualifying metadata.

A Web service is the access point to the real service: its quality determines the quality of access, but might be independent of the quality of service (e.g. automatic translation or flight reservation). Therefore, an iterative selection process (similarly to iterative query refinement in information retrieval) guides the selection of the best service in INFRAWEB. This is based on a simple two-phase workflow agreement (or business logic) between the web services and the middleware layer.

The SUA component of INFRAWEB serves as a generic middleware for deployment of Semantic Web Services. It provides support for the usual steps of goal construction, discovery, selection and execution of Semantic Web Services. This support is achieved through a neutral interface, which hides the complexities of Semantic Web Services, therefore applications can be easily adapted to it, and also it can be equivalently used in variants of Semantic Web Services, such as WSMO and OWL-S. SUA also features an iterative selection refinement process for finding not only the suitable web services, but also the best service offers for users' goals.

**SWS deployment layer** The SWS deployment layer consist of SWS executor that is split up into three main components, namely, the Communication Manager, Choreography Engine and the Invoker [Polleres, 2005] and QoS broker. At present it is defined that the executor should mainly interact with the distributed registry and the SUA components.

---

## Conclusion

---

One of the goals of the INFRAWEB project is the development of a SWS full-life-cycle software toolset for creating and maintaining Semantic Web Services with specific application support. An important part of INFRAWEB is a Semantic Web Unit (SWU) – a collaboration platform and interoperable middleware for ontology-based handling and maintaining of SWS. The SWU provides knowledge about a specific domain and relies on ontologies to structure and exchange this knowledge.

INFRAWEB Designer and Composer are sub-modules of SWU responsible for creating Semantic Web Services using Case-Based Reasoning approach to fulfil decision support demands.

The architecture of both modules is based on such general principles as:

- service-oriented architecture with bottom-up approach for semi-automatic constructing of semantic web services;
- system driven syntactic consistent and completeness checking;
- past experience utilizing.

The SUA unit within INFRAWEB acts as a middleware layer for user applications, connecting these applications to the functionality available in the form of Semantic Web Services. These software toolsets are developing for creating and maintaining of full-life-cycle Semantic Web Services with specific application support.

---

## Bibliography

---

- [Nern, 2004] H Joachim Nern, G. Agre, T. Atanasova, J. Saarela. System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEB II. In: WSEAS TRANS. on INFORMATION SCIENCE and APPLICATIONS, 1, Vol. 1, 286-291, 2004.
- [Nern, 2005a] H Joachim Nern, A Dziech, E Tacheva, Fuzzy Concept Sets (FCS) applied to semantic organizational memories within the Semantic Web Service designing and composing cycle, In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005.
- [Zadeh, 2002] Lotfi A. Zadeh. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. In: Journal of Statistical Planning and Inference 105, 233-264, 2002.
- [Westkaemper, 2005] T Westkaemper, J Saarela, H Joachim Nern, Semantic Information routing as a pre-process for Semantic Web Service generation - SIR & OM, In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005.

- [Atanasova, 2005] Tatiana Atanasova, Gennady Agre, H Joachim Nern, "INFRAWEBs Semantic Web Unit for design and composition of Semantic Web Services INFRAWEBs approach", In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005.
- [Nern, 2005b] H Joachim Nern, A Dziech, T Atanasova, Applying Clustering and Classification Methods to distributed Decision Making in Semantic Web Services Maintaining and Designing Cycles, EUROMEDIA 2005, In: Proc. Workshop for "Semantic Web Applications", IRIT, Université Paul Sabatier, Toulouse, France, April 11-13, 2005.
- [Agre, 2005] Gennady Agre, Tatiana Atanasova, H Joachim Nern, "Case Based Designer and Composer", In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005.
- [Kovacs, 2005] L Kovacs, A Micsik, "The SUA-Architecture within the Semantic Web Service Discovery and selection process", In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005.
- [Polleres, 2005] A Polleres, J Scicluna, "Semantic Web Execution for WSMO based choreographies", In: Proc. 1st Workshop for "Semantic Web Applications" at the EUROMEDIA 2005, IRIT, Université Paul Sabatier, Toulouse, France, April 2005

---

### Authors' Information

---

**Joachim Nern** – Scientific coordinator of INFRAWEBs project; big7.net GmbH & Aspasia Knowledge Systems Germany, e-mail: [nern@aspasia-systems.de](mailto:nern@aspasia-systems.de)

**Tatiana Atanasova** – Institute of Information Technologies, Acad. G. Bonchev 2, 1113 Sofia, Bulgaria, e-mail: [atanasova@iinf.bas.bg](mailto:atanasova@iinf.bas.bg)

**Gennady Agre** – Institute of Information Technologies, Acad. G. Bonchev 29-A, 1113 Sofia, Bulgaria, e-mail: [agre@iinf.bas.bg](mailto:agre@iinf.bas.bg)

**András Micsik** – MTA SZTAKI, H-1111 Budapest XI. Lagymányosi u. 11, Hungary, e-mail: [micsik@sztaki.hu](mailto:micsik@sztaki.hu)

**László Kovács** – MTA SZTAKI, H-1111 Budapest XI. Lagymányosi u. 11, Hungary, e-mail: [Laszlo.kovacs@sztaki.hu](mailto:Laszlo.kovacs@sztaki.hu)

**Janne Saarela** – Profium Ltd, 02600 Espoo, Finland, e-mail: [janne.saarela@profium.com](mailto:janne.saarela@profium.com)

**Timo Westkamper** – Profium Ltd, 02600 Espoo, Finland, e-mail: [timo.westkamper@profium.com](mailto:timo.westkamper@profium.com)

## ADJUSTING WSMO API REFERENCE IMPLEMENTATION TO SUPPORT MORE RELIABLE ENTITY PERSISTENCE<sup>1</sup>

Ivo Marinchev

**Abstract:** *In the presented paper we scrutinize the persistence facilities provided by the WSMO API reference implementation. It is shown that its current file data-store persistence is not very reliable by design. Our ultimate goal is to explore the possibilities of extending the current persistence implementation (as an easy short-run solution) and implementing a different persistent package from scratch (possible long-run solution) that is more reliable and useful. In order to avoid "reinventing the wheel", we decided to use relational database management system to store WSMO API internal object model. It is shown later that the first task can be easily achieved although in not very elegant way, but we think that the later one requires some changes in the WSMO API to smooth out some inconsistencies in the WSMO API specification in respect to other widely used Java technologies and frameworks.*

**Keywords:** *Semantic Web Services, Web Service Modelling Ontology (WSMO), WSMO API, WSMO4J.*

---

<sup>1</sup> The research has been partially supported by INFRAWEBs - IST FP62003/IST/2.3.2.3 Research Project No. 511723 and "Technologies of the Information Society for Knowledge Processing and Management" - IIT-BAS Research Project No. 010061.

---

## Introduction

---

Web services are defining a new paradigm for the Web in which a network of computer programs becomes the consumer of information. However, Web service technologies only describe the syntactical aspects of a Web service and, therefore, only provide a set of rigid services that cannot be adapted to a changing environment without human intervention. Realization of the full potential of the Web services and associated service oriented architecture requires further technological advances in the areas of service interoperation, service discovery, service composition and orchestration. A possible solution to all these problems is likely to be provided by converting Web services to *Semantic Web* services. *Semantic Web* services are “self-contained, self-describing, semantically marked-up software resources that can be published, discovered, composed and executed across the Web in a task driven semi-automatic way” [Arroyo et al, 2004].

There are two major initiatives aiming at developing world-wide standard for the semantic description of Web services. The first one is OWL-S [OWL-S, 2004], a collaborative effort by BBN Technologies, Carnegie Mellon University, Nokia, Stanford University, SRI International and Yale University. The second one is Web Service Modelling Ontology (WSMO) [Roman et al, 2004], a European initiative intending to create an ontology for describing various aspects related to Semantic Web Services and to solve the integration problem.

As part of the later initiative the WSMO API specification and reference implementation [WSMO4J] has been developed. WSMO4J is an API and a reference implementation for building Semantic Web Services applications compliant with the Web Service Modeling Ontology. WSMO4J is compliant with the WSMO v1.0 specification [WSMO v1.0] (20 Sep 2004). At the time of this writing the WSMO API reference implementation is version 0.3 (alpha), that means that it is far from completed product and is subject to changes without prior notice. Nevertheless the API is incomplete as part of our work on the INFRAWEB project [Nern et al, 2004] we have to utilize the WSMO4J package as it is the only available working implementation of the WSMO specifications.

---

## Current State of the Art

---

At the time of this writing the reference implementation of the WSMO API [WSMO4J] can export its internal data model to a set of binary files organized in a bundle of directories that correspond to the major entity (entities that are identifiable according to WSMO API terms) types. Every identifiable entity is saved as a separate file that contains the serialized entity identifier, then the Java object that represents the entity and at the end several lists of identifiers corresponding to the different groups of entities that are subordinates of the current entity. The order in which these lists are serialized to the output file (stream) is implementation specific and is implemented by several internal classes named entity-type processors. Every entity type has separated processor class that serializes/deserializes the corresponding objects to/from their persistent state. The subordinate entities are stored in the same way in separate files and so on. Fig. 1 shows the directories created by the file data-store. In practice, this simple solution appears to be very unreliable and a relatively small problem may incur enormous data-losses. The reader also has to keep in mind that this storage mechanism is intended to be used for processing ontologies. And all of the ontologies that are applicable to real-world problem tend to be extremely large.

This ad hoc solution is simple and does not require any third party programs/libraries but it has several major drawbacks that prevent it to be used in a production system. Some of them are implementation problems and can be easily circumvented but others require rather sophisticated solutions to general purpose problems.

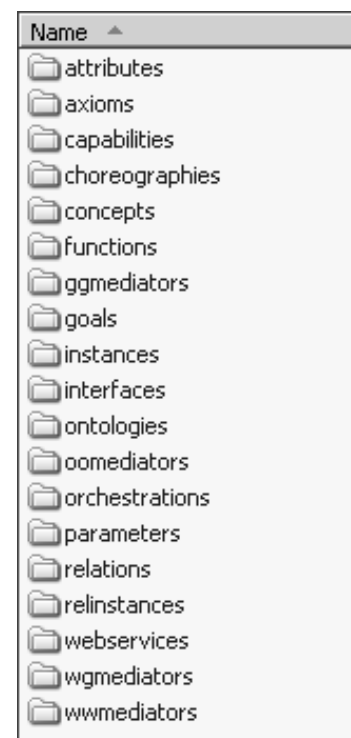


Fig. 1: Directories created by the “FileDataStore” class of the reference implementation.



---

Below we enumerate some of the problems:

1. If a certain entity is loaded, all its dependent entities are loaded as well no matter whether they are needed or not.
2. If any object has to be changed all its subordinate objects are overwritten again no matter if they are changed or not. Moreover because the implementation of the ObjectOutputStream in Java serializes objects by reachability (when an object is serialized all objects that are referenced by it are serialized as well) a certain object is serialized as many times as the number of objects that refers to it in *direct and indirect way!* *Such persistence scheme brings enormous excessive overhead in the serialization of large object graphs and the worst is that the overhead increases exponentially with the size of the object graph.* A possible solution to the later problem is all objects to implement Externalizable interface in order to control their serialization process but even it will not remove all unneeded read/write operations.
3. In the current implementation, the file names correspond to the entity identifiers. These identifiers may become rather long. This is especially important as many identifiers are created from URLs and at the same time, many of the entities have the identifiers that extend the identifier of its parent entity (for example axioms that are part of the ontology). The implementation makes the file names additionally longer by encoding some special characters that may be prohibited by certain file systems (for example / is encoded with .fslash., : with .colon., \* with .star.). At the same time most of the file systems do not allow file names with more than 255 characters.
4. The store operations are not atomic. Thus, there is no guarantee that the data will remain consistent and the original object graph can be recreated from its serialized state. For example if an exception occurs when the data is being saved, the operation is terminated and the on-disk structures remain in an unpredictable and undeterminable state – the user neither can fix them, nor can turn them to their state before the last operation occur (roll back the last operation).
5. The store implementation is not thread-safe due to the usage of fields to transfer data between methods. But even that one can instantiate several different data-store objects they can not work on the same store simultaneously because of the lack of any locking or synchronizations.

In short, if we use database terminology, the current implementation is very far from being ACID<sup>1</sup> compliant. It is obvious that any of the above issues can be solved but the solutions are usually very sophisticated, and one ultimately will implement complete transactional database storage engine in order to solve all of them.

---

### Using Relational Database as a Data Store

---

The first improvement that we implemented was to move the persistent data to the relational database by just replacing the file data-store directories with the database tables. The tables consist of two columns: the first one for the entity identifier, and the second column of type BLOB (Binary Large Object) that stores the serialized Java objects. This extension was relatively easy to be implemented. We changed several private methods that deal with the file names and entity types (getFileNameFor, getEntityType, etc.) to work with the database tables and records, and then we changed all of methods that serialize and deserialize data to store/load it to the corresponding BLOB fields instead of using file output and input streams. In order to be able to use the transaction facilities provided by all modern relational database management systems (RDBMS) we use a single database connection that is initiated at the beginning of the store process and get committed at its end (or rolled back in case of exceptional circumstances).

Even with these simple modifications, we get several important advantages:

1. We get atomic changes – all of the changes are written or all are discarded at once.
2. No data is lost in case the storing gets terminated - not only by checked exception but even if the whole process is terminated by the unchecked one.
3. The store may be physically located on the remote system.
4. Several different client processes can use the store simultaneously.

---

<sup>1</sup> ACID – Atomicity (states that database modifications must follow an “all or nothing” rule), Consistency (states that only valid data will be written to the database), Isolation (requires that multiple transactions occurring at the same time not impact each other’s execution), Durability (ensures that any transaction committed to the database will not be lost).

5. The store may use the back-up, replication, and clustering facilities provided by the underlying RDBMS.
  6. The store uses the data caching provided by the database.
  7. The database can be changed at will if one does not use proprietary database extensions.
- 

### **Avoiding Identity-Lists Serialization**

---

The next logical consequent step is to start removing object serializations. As it was discussed earlier when a certain entity is persisted the file data-store in the reference implementation serializes first the entity identifier, then the entity object and at the end several lists of the identifiers of the entities that depend on the current one in the entity hierarchy. This last step is entity type specific and is implemented by a specialized entity processors for different type of entities that take care of saving, loading the lists (Vectors in Java terms) of identifiers.

Using the information from these entity processor classes, we created a separate table columns that hold the lists of identities of a given type. For example, for the “capability processor”, we created columns for Assumptions list, Pre-Conditions list, Post-Conditions lists, and Effects list. Thus, the content of the database tables gets more human readable and it is easier to debug potential problems, but we have to emphasize that the database is still not even in the first normal form (1NF) - it requires all table columns to be atomic.

---

### **Utilizing Object-Relational Frameworks**

---

The newly created columns in fact represent the relationships between the entities represented by the corresponding table rows and their dependent entities. That is why we can remove these columns and replace them with foreign key columns in the dependent tables for one-to-many relationships and with relationship tables for the many-to-many relationships. Dealing with one-to-many relationships with hand-written code is boring but not a complicated task. But the many-to-many relationships can become really problematic to be manipulated as they use additional (relationship) tables and the WSMO object model even have many-to-many reflexive relationships (for example the one between concepts and sub-concepts of the ontology). These facts imply that the programming code needed to deal with all these “housekeeping” activities will be much more than the code that implements the actual business logic.

At this point one can realize that the required changes to the original reference implementation become rather complex and in fact we start “reinventing the wheel” that is already created by others. So, the wise approach to this problem is to use some object-relational mapping frameworks to do the work for us. There are a lot of such frameworks available (for example Java Data Objects [JDO] implementations, Hibernate [Hibernate], Oracle TopLink [TopLink], and others) and many of them are open-source and free even for commercial use. The common feature of all these frameworks is that they use XML configuration files to specify how the objects, fields and relationships (object model) are mapped to the corresponding database tables and columns (relational model). The basic idea behind these mapping files is to keep the object model and the relational model loosely coupled so that the two models can be changed independently. Utilizing such framework provides other useful features:

1. Automatic generation of database queries;
2. Loading/saving/updating the complete object graph with a single method call;
3. Lazy-loading (or on-demand loading)<sup>1</sup>;
4. Tracking the user changes and updating just the changed fields;
5. Support for many different RDBMS;
6. Object caching - even distributed caching is possible;
7. Other specific features.

For our purpose the lazy-loading is extremely useful because if one wants to load and change a certain entity it does not need to load and save the complete sub-graph that originates from this entity.

---

<sup>1</sup> The framework loads the expensive (in memory footprint and construction time) object fields and referenced objects just before they are accessed by the client program. This feature is usually very flexible and can be configured in the mapping files on a field level.

Unfortunately, it appears that several significant issues arise in any attempt of integration between object-relational mapping framework and the current version of the WSMO API and its reference implementation. These issues are discussed in the next section. At the end of it, we represent one possible solution of the problem and why we think the proposed changes are appropriate.

---

### Problems with the Current Version of WSMO API and its Reference Implementation

---

The most serious problem concerning the applicability of the OR mapping framework with the WSMO API (and its reference implementation) is that the object-relational frameworks work with JavaBeans classes/objects. We do not know why WSMO API was specified and implemented in its current form, but the fact is that all of the entity classes in it deviate from the JavaBeans specification [JavaBeans]. Specifically they lack the properly named accessor and mutator methods for the non-primitive types. At the same time, all methods for accessing non-primitive types are named as listXXX. We do not know why such naming scheme has been selected but we think that it is even not very intuitive. The worst is that at the same time listXXX methods return value is of type `java.util.Set`. In fact, the following issues appear:

1. The names of the property accessor methods are misleading for the user and deviate from other well-known framework and the JavaBean specification.
2. The semantics of the Set and List data types are significantly different as the list is an ordered collection of elements. More over unlike sets, lists typically allow duplicate elements. More formally, lists typically allow pairs of elements  $e_1$  and  $e_2$  such that  $e_1.equals(e_2)$ , and they typically allow multiple null elements if they allow null elements at all. So the words list and set is not very appropriate to be used interchangeably.
3. Using "un-typed" return types in the listXXX methods in the otherwise very strongly typed specification is somewhat strange decision.

It is true that we can overcome the first problem by sub-classing all needed entity classes of the reference implementation and turn them to regular JavaBeans by adding the missing accessor and mutator methods and then create mappings for the newly introduced classes. But we do not want any consequent version of the reference implementation to break our "extension", or to require conversions of the database schema. So, this solution does not seem appropriate in the long-run.

At the end we will express our inner conviction that the persistence package has to be as loosely coupled as possible to the rest of the implementation, and to be written as much as possible against the specification not against the implementation as it is now. In the confirmation of the later we propose the WSMO API specification to be changed in the following way:

1. Add the missing get/set methods to the entity interfaces to turn the implementation classes in correct JavaBeans.
2. Introduce type-safe sets for any entity type that is needed and return them in the corresponding property accessor methods instead of `java.util.Set`.

---

### Conclusion

---

As a conclusion we want to point out that although it is in its early stage of development and the fact that it is or still may be immature in some of its parts, the WSMO4J is sound enough to be used as a development tool in the research projects, and facilitates researchers in the early adoption of the WSMO related technologies. We hope that the WSMO API working group will take into account our remarks and suggestions and even they are rejected they will contribute in some way in the future improvements of the specification and/or implementation.

---

### Bibliography

---

- [Arroyo et al, 2004] Arroyo, S., Lara, R., Gomez, J. M., Bereka, D., Ding, Y., Fensel, D. Semantic Aspects of Web Services. In: Practical Handbook of Internet Computing. Munindar P. (Editor). Chapman Hall and CRC Press, Baton Rouge, 2004
- [JDO] <http://java.sun.com/products/jdo/>
- [Hibernate] <http://www.hibernate.org>
- [JavaBeans] <http://java.sun.com/products/javabeans/>

- [Nern et al. 2004] H.-Joachim Nern, G. Agre, T. Atanansova, J. Saarela. System Framework for Generating Open Development Platforms for Web-Service Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems - INFRAWEB II. *WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS*, ISSN 1790-0832, Issue 1, Volume 1, July 2004, 286-291.
- [OWL-S, 2004] The OWL Services Coalition: OWL-S: Semantic Markup for Web Services, *version 1.0*; available at <http://www.daml.org/services/owl-s/1.0/owl-s.pdf>
- [Roman et al. 2004] D. Roman, U. Keller, H. Lausen (eds.): Web Service Modeling Ontology (WSMO), version 0.1; available at: <http://nextwebgeneration.com/projects/wsmo/2004/d4/d4.1/v01/index.html>
- [TopLink] <http://www.oracle.com/technology/products/ias/toplink/>
- [WSMO v1.0] <http://www.wsmo.org/2004/d2/v1.0/20040920/>
- [WSMO4J] <http://wsmo4j.sourceforge.net>
- 

### Author's Information

---

**Ivo Marinchev** – Institute of Information Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 29A, Sofia-1113, Bulgaria; e-mail: [ivo@iinf.bas.bg](mailto:ivo@iinf.bas.bg)

## INFRAWEB CAPABILITY EDITOR – A GRAPHICAL ONTOLOGY-DRIVEN TOOL FOR CREATING CAPABILITIES OF SEMANTIC WEB SERVICES

**Gennady Agre, Peter Kormushev, Ivan Dilov**

**Extended Abstract:** The current INFRAWEB European research project aims at developing ICT framework enabling software and service providers to generate and establish open and extensible development platforms for Web Service applications. One of the concrete project objectives is developing a full-life-cycle software toolset for creating and maintaining Semantic Web Services (SWSs) supporting specific applications based on Web Service Modelling Ontology (WSMO) framework. SWSs are self-contained, self-describing, semantically marked-up software resources that can be published, discovered, composed and executed across the Web in a task driven semi-automatic way. A main part of WSMO-based SWS is service capability – a declarative description of Web service functionality. A formal syntax and semantics for such a description is provided by Web Service Modeling Language (WSML), which is based on different logical formalisms, namely, Description Logics, First-Order Logic and Logic Programming. A WSML description of a Web service capability is represented as a set of complex logical expressions (axioms). The paper describes a specialized user-friendly tool for constructing and editing WSMO-based SWS capabilities. Since the users of that tool are assumed to be SWS providers, which are not the specialists in first-order logic, it is proposed a graphical way for constructing and editing the axioms abstracting away as much as possible from a concrete syntax of logical language used for implementing them.

Our analysis has shown that the main problems arising during axiom creating are associated with using of correct names of concepts, attributes, relations and parameters as well as their types rather than with expressing logical dependences between axiom parts. So the process of constructing logical expressions in the tool is ontology-driven, which means that in each step of this process the user may select only such elements of existing ontologies that are consistent with already constructed part of the axiom. From this point of view the created axiom is always semantically consistent with ontologies used for its construction. After discussing the main design principles of the Editor, its functional architecture is briefly presented. The tool is implemented in Eclipse Graphical Environment Framework and Eclipse Rich Client Platform.

---

### Authors' Information

---

**Gennady Agre** – Institute of Information Technologies – BAS, e-mail: [agre@iinf.bas.bg](mailto:agre@iinf.bas.bg)

**Peter Kormushev** – Sofia University St. Kliment Ohridski, e-mail: [pkormushev@ppartner.com](mailto:pkormushev@ppartner.com)

**Ivan Dilov** – Sofia University St. Kliment Ohridski, e-mail: [idilov@ppartner.com](mailto:idilov@ppartner.com)

## INDEX OF AUTHORS

Adil V. Timofeev	108	Juan Castellanos Peñuela	58
Adriana Toni	18	Katalina Grigorova	181
Alexander Elkin	169	László Kovács	217
Alexander Fish	51	Lev Grigoriev	210
Alexander Koshchy	169	Levon Aslanyan	7,12
Alexander Kuzemin	169,204	Ludmila Kirichenko	124
Alexandr Karpukhin	102	Luis Fernández	115
Alexey Zablotzky	210	Luis Fernando de Mingo López	66
Anatoli Nachev	95	Martin P. Mintchev	39
András Micsik	217	Mercedes Perez-Castellanos	79
Andrey D. Danilov	190	Micael Gallego-Carrillo	110
Andrey V. Stolyarenko	22	Mikhail Bondarenko	102
Asanbek Torojev	204	Milena Dobрева	149
Boris Elkin	169	Nadezhda N. Kiselyova	22
Carmen Torres	79,87	Natalia Nosovich	171
Carolina Gallardo	71	Nikola Ikononov	149
Charles Newton Price	39	Nikolay Korolev	171
Cristina Hernández de la Sota	58	Orly Yadid-Pecht	51
David T. Westwick	39	Peter Kormushev	228
Dimitrina Polimirova–Nickolova	130	Plamenka Hristova	181
Dmitriy P. Murat	22	Rafael Gonzalo Molina	58
Dmitry Kudryavtsev	210	Renato J. de Sobral Cintra	39
Elena Bulavina	204	Rosalía Peña	115
Elena Castiñeira	87	Sergey Georgiev	135
Elena I. Bolshakova	155	Sergey Kiprushkin	171
Eugene Nickolov	130	Sergey Kurskov	171
Eugenio Santos	71	Soto Montalvo-Herranz	110
Evgenij A. Eremin	165	Stoyan Poryazov	141
Francisco Gisbert	66	Susana Cubillo	87
Galina Atanasova	181	Svetlana Chumachenko	124
Galina Bogdanova	33	Svetlana Roshka	102
Ganna Molodykh	160	Tatiana Atanasova	217
Gennady Agre	217,228	Timo Westkaemper	217
Georgi Stoilov	143	Todorka Kovacheva	200
Grigorij Chetverikov	102	Tsvetanka Georgieva	33
Hasmik Sahakyan	12	Valentín Palencia Alejandro	58
Ilya V. Prokoshev	22	Valentin V. Khorbenko	22
Irina Radeva	189	Valentina Kislova	210
Ivailo Petkov	135	Valeriy Bykov	160
Ivan Dilov	228	Valery Kornyshko	27
Iván García-Alcaide	110	Victor A. Dudarev	22,27
Ivan Popchev	189	Víctor Giménez-Martínez	79
Ivo Marinchev	223	Victor S. Zemskov	22
Janne Saarela	217	Victoria Zarzosa	87
Jennifer Q. Trelewicz	181	Vyacheslav Liashenko	204
Jesús Cardeñosa	71	Yuriy Zhook	160
Joachim Nern	217	Zhanna Deyneko	102
José Joaquín Erviti Anaut	18,79		



**International Journal  
“Information Theories and Applications”**



***International Conferences  
on Information Theories and Applications  
supported and published by IJ ITA***



***Knowledge,  
Dialogue,  
Solutions***



***General  
Information  
Theory***

***i.tech***

***Information Research,  
Applications  
and Education***

***Bi***

***Business  
Informatics***

***MULTIMEDIA  
SEMANTICS***

***Multimedia  
Semantics***



***Digitisation  
of Cultural Heritage***

***General sponsor:***

**FOI BULGARIA**



© 2005 Markov, Ivanova, Mitov

***For more information:***

**[www.foibg.com](http://www.foibg.com)**