

**INFORMATION SCIENCE  
&  
COMPUTING**

---

International Book Series

Number 9

---

# **Intelligent Processing**

---

Supplement to  
International Journal "Information Technologies and Knowledge" Volume 3 / 2009

---

**ITHEA  
SOFIA, 2009**

Krassimir Markov, Peter Stanchev, Krassimira Ivanova, Ilia Mitov (ed.)

Intelligent Processing

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 9

Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3 / 2009

Institute of Information Theories and Applications FOI ITHEA

Sofia, Bulgaria, 2009

This issue contains a collection of papers in the field of Natural Language Processing, Neural Nets, Evolutionary and Genetic Algorithms.

Papers are peer reviewed and are selected from the several International Conferences, which were part of the Joint International Events of Informatics "ITA 2009" – summer session, Varna, Bulgaria.

**International Book Series "INFORMATION SCIENCE & COMPUTING", Number 9**  
**Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3, 2009**

Edited by **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria,  
in collaboration with

- **V.M.Glushkov Institute of Cybernetics of NAS**, Ukraine,
- **Institute of Mathematics and Informatics, BAS**, Bulgaria,
- **Institute of Information Technologies, BAS**, Bulgaria.

Publisher: Institute of Information Theories and Applications FOI ITHEA, Sofia, 1000, P.O.B. 775, Bulgaria.  
Издател: Институт по информационни теории и приложения ФОИ ИТЕА, София, 1000, п.к. 775, България  
[www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com), e-mail: [info@foibg.com](mailto:info@foibg.com)

General Sponsor: **Consortium FOI Bulgaria** ([www.foibg.com](http://www.foibg.com)).

Printed in Bulgaria

**Copyright © 2009 All rights reserved**

- © 2009 Institute of Information Theories and Applications FOI ITHEA - Publisher
- © 2009 Krassimir Markov, Peter Stanchev, Krassimira Ivanova, Ilia Mitov – Editors
- © 2009 For all authors in the issue.

**ISSN 1313-0455 (printed)**

**ISSN 1313-048X (online)**

**ISSN 1313-0501 (CD/DVD)**

---

## PREFACE

The scope of the International Book Series "Information Science and Computing" (**IBS ISC**) covers the area of Informatics and Computer Science. It is aimed to support growing collaboration between scientists from all over the world. IBS ISC is official publisher of the works of the members of the ITHEA International Scientific Society.

The official languages of the IBS ISC are English and Russian.

IBS ISC welcomes scientific papers and books connected with any information theory or its application.

IBS ISC rules for preparing the manuscripts are compulsory.

The rules for the papers and books for IBS ISC are given on [www.foibg.com/ibsisic](http://www.foibg.com/ibsisic) .

The camera-ready copies of the papers should be received by ITHEA Submission System <http://ita.ithea.org> .

The camera-ready copies of the books should be received by e-mail: [info@foibg.com](mailto:info@foibg.com) .

Responsibility for papers and books published in IBS ISC belongs to authors.

This issue contains a collection of papers in the field of Natural Language Processing, Neural Nets, Evolutionary and Genetic Algorithms.

Papers are peer reviewed and are selected from the several International Conferences, which were part of the Joint International Events of Informatics "ITA 2009" – summer session, Varna, Bulgaria.

ITA 2009 has been organized by

ITHEA International Scientific Society

in collaboration with:

- Institute of Information Theories and Applications FOI ITHEA
- International Journal "Information Theories and Applications"
- International Journal "Information Technologies and Knowledge"
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Association for Development of the Information Society (Bulgaria)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Information Technologies, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- Taras Shevchenko National University of Kiev (Ukraine)
- Dorodnicyn Computing Centre of the Russian Academy of Sciences
- Universidad Politecnica de Madrid (Spain)
- BenGurion University (Israel)
- Rzeszow University of Technology (Poland)
- University of Calgary (Canada)
- University of Hasselt (Belgium)
- Kharkiv National University of Radio Electronics (Ukraine)
- Astrakhan State Technical University (Russia)
- Varna Free University "Chernorizets Hrabar" (Bulgaria)
- National Laboratory of Computer Virology, BAS (Bulgaria)
- Uzhgorod National University (Ukraine)

The main ITA 2009 events were:

<b>KDS</b>	XVth International Conference "Knowledge - Dialogue – Solution"
<b>i.Tech</b>	Seventh International Conference "Information Research and Applications"
<b>MeL</b>	Fourth International Conference "Modern (e-) Learning"
<b>INFOS</b>	Second International Conference "Intelligent Information and Engineering Systems"
<b>CFDM</b>	International Conference "Classification, Forecasting, Data Mining"
<b>GIT</b>	Seventh International Workshop on General Information Theory
<b>ISSI</b>	Third International Summer School on Informatics

More information about ITA 2009 International Conferences is given at the [www.ithea.org](http://www.ithea.org) .

The great success of ITHEA International Journals, International Book Series and International Conferences belongs to the whole of the ITHEA International Scientific Society.

We express our thanks to all authors, editors and collaborators who had developed and supported the International Book Series "Information Science and Computing".

General Sponsor of IBS ISC is the **Consortium FOI Bulgaria** ([www.foibg.com](http://www.foibg.com)).

*Sofia, June 2009*

*Kr. Markov, P. Stanchev, Kr. Ivanova, I. Mitov*

## TABLE OF CONTENTS

<i>Preface</i> .....	3
<i>Table of Contents</i> .....	5
<i>Index of Authors</i> .....	7

### **Natural Language Processing**

Mobile Search and Advertising <i>Vladimir Lovitskii, Colin McCaffery, Michael Thrasher, David Traynor, Peter Wright</i> .....	9
Mobile Election <i>Elena Long, Vladimir Lovitskii, Michael Thrasher, David Traynor</i> .....	19
Computer Support of Semantic Text Analysis of a Technical Specification on Designing Software <i>Alla Zaboleeva-Zotova, Yulia Orlova</i> .....	29
К анализу естественно-языковых объектов <i>Александр Палагин, Сергей Кривый, Виталий Величко, Николай Петренко</i> .....	36
Обработка предложений естественного языка с использованием словарей и частоты появления слов <i>Александр Палагин, Сергей Кривый, Дмитрий Бибиков</i> .....	44
Определение понятия «смысл» через онтологию. Семантический анализ текстов естественного языка <i>Леонид Святогор, Виктор Гладун</i> .....	53

### **Neural Nets**

Self-learning Fuzzy Spiking Neural Network as a Nonlinear Pulse-position Threshold Detection Dynamic System Based on Second-order Critically Damped Response Units <i>Yevgeniy Bodyanskiy, Artem Dolotov, Iryna Pliss</i> .....	63
Performance Comparison of MATLAB and Neuro Solution Software on Estimation of Fuel Economy by Using Artificial Neural Network <i>Hakan Serhad Soyhan, Mehmet Emre Kilic, Burak Gokalp, Imdat Taymaz</i> .....	71
The Usage of Neural Networks for the Medical Diagnosis <i>Kateryna Malyshevskaya</i> .....	77
Инструментальная среда для исследования эволюционных стратегий с использованием нейросетевых метамоделей <i>Павел Афонин</i> .....	81
Применение нейронной сети Хемминга и нечеткой логики к обнаружению краев объектов на изображениях в оттенках серого <i>Николай Мурга</i> .....	87
Отображение и вывод по аналогии на основе нейросетевых распределенных представлений <i>Сергей Слипченко, Дмитрий Рачковский</i> .....	95

---



---

## **Evolutionary and Genetic Algorithms**

The Production Scheduling in Assembly System with Evolutionary Algorithm <i>Galina Setlak</i> .....	103
The Cascade Neo-fuzzy Architecture and its Online Learning Algorithm <i>Yevgeniy Bodyanskiy, Yevgen Viktorov</i> .....	110
Analysis of P-Systems under a Multiagent Systems Perspective <i>Alberto Arteta, Angel Goñi, Juan Castellanos</i> .....	117
Implementation of Genetic Algorithms for Transit Points Arrangement <i>Dmitry Panchenko, Maxim Shcherbakov</i> .....	129
Self Evolving Character Recognition Using Genetic Operators <i>Shashank Mathur</i> .....	132
Generating More Boundary Elements of Subset Projections <i>Hasmik Sahakyan, Levon Aslanyan</i> .....	139
METASPEED: Metadata ExTraction for Automatised SPEcifications of E-Documents <i>About Project D 002-308 (Bulgarian NSF)</i> .....	145

---



---

## INDEX OF AUTHORS

Alberto Arteta	117	Павел Афонин	81
Levon Aslanyan	139	Дмитрий Бибииков	44
Yevgeniy Bodyanskiy	63, 110	Виталий Величко	36
Juan Castellanos	117	Виктор Гладун	53
Artem Dolotov	63	Сергей Крывый	36, 44
Burak Gokalp	71	Николай Мурга	87
Angel Goñi	117	Александр Палагин	36, 44
Mehmet Kilic	71	Николай Петренко	36
Elena Long	19	Дмитрий Рачковский	95
Vladimir Lovitskii	9, 19	Леонид Святогор	53
Kateryna Malyshevskaya	77	Сергей Слипченко	95
Shashank Mathur	132		
Colin McCaffery	9		
Yulia Orlova	29		
Dmitry Panchenko	129		
Iryna Pliss	63		
Hasmik Sahakyan	139		
Galina Setlak	103		
Maxim Shcherbakov	129		
Hakan Soyhan	71		
Imdat Taymaz	71		
Michael Thrasher	9, 19		
David Traynor	9, 19		
Yevgen Viktorov	110		
Peter Wright	9		
Alla Zabooleeva-Zotova	29		





---



---

## Natural Language Processing

---



---

### MOBILE SEARCH AND ADVERTISING

**Vladimir Lovitskii, Colin McCaffery, Michael Thrasher, David Traynor, Peter Wright**

**Abstract:** Mobile advertising is a rapidly growing sector providing brands and marketing agencies the opportunity to connect with consumers beyond traditional and digital media and instead communicate directly on their mobile phones. Mobile advertising will be intrinsically linked with mobile search, which has transported from the internet to the mobile and is identified as an area of potential growth. The result of mobile searching show that as a general rule such search result exceed 160 characters; the dialog is required to deliver the relevant portion of a response to the mobile user. In this paper we focus initially on mobile search and mobile advert creation, and later the mechanism of interaction between the user's request, the result of searching, advertising and dialog.

**Keywords:** mobile text messages, mobile search, mobile advertising, question-answering system

**ACM Classification Keywords:** I.2 Artificial intelligence: I.2.7 Natural Language Processing: Text analysis.

**Conference:** The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009

---

#### Introduction

---

This paper considers the results of our recent research in the areas of text data mining and the natural language processing [1-7] when restricted by mobile phone text-based SMS messaging. In our previous papers [6,7] the Question-Answering Mobile Engine (QAMEN) has been discussed. During internet searching QAMEN converts web pages to a simplified format that is compatible with handheld devices. Moreover, QAMEN frees users to have an expensive mobile phone with a web browser. Internet connections from mobile devices continue to remain expensive and there is little prospect to an immediate decrease in pricing structure. QAMEN is based on industry-standard SMS messaging technology and thus works with any mobile in any GSM network. Hence QAMEN is useful for people on the move and probably unable to access a PC.

**Worldwide Mobile Search Advertising Spending, by Region, 2007-2012 (millions)**

	2007	2008	2009	2010	2011	2012
US	\$34.5	\$107.4	\$241.8	\$530.5	\$910.2	\$1,484.2
Asia-Pacific	\$26.0	\$72.0	\$189.9	\$372.8	\$732.4	\$1,160.0
Western Europe	\$18.4	\$52.0	\$140.5	\$339.7	\$614.1	\$968.2
Rest of World	\$4.4	\$12.4	\$24.4	\$47.0	\$88.6	\$160.9
<b>Worldwide</b>	<b>\$83.3</b>	<b>\$243.7</b>	<b>\$596.6</b>	<b>\$1,289.9</b>	<b>\$2,345.2</b>	<b>\$3,773.2</b>

The growth of mobile-search-related advertising is strong across many regions [8] (see the Table). Mobile search will eventually move away from the generic style of searches seen on the fixed internet to more personal services. In these circumstances local search will become increasingly important. Through these local

searches **location-based services and advertising will be able to gain vital revenues** as mobile subscribers using search will reveal both their geographical location and the items being searched for.

Mobile search and mobile advertising should grow significantly because mobile phones have several advantages over PCs when accessing the Internet [9]:

- Mobile phones are always on, always available, and always “connected”. Subscribers can gain access to information anywhere; at home, in the office, at a restaurant, or from the car.
- There are currently 3 billion mobile phones worldwide compared with less than 1 billion PCs.
- Ability to immediately connect people to phone numbers, since they already have a device in hand.
- When conducting mobile search, users are in an atmosphere, situation and environment more likely to result in a purchase. Most of the time they are out of the home.

But mobile phones have some limitations when compared to PCs:

- PCs have relatively large screens that can show more detailed information and numerous search results. By contrast, mobile phones have smaller screens with limited real estate.
- PCs have large, comfortable keyboards with easy-to-use pointing tools including mouse, trackball, or touchpad. Mobile phones have compact number pads, commonly with arrow keys limiting user-navigation to up, down, left and right.
- PCs and the Internet are relatively homogenous, using consistent colour displays, screen sizes, browsers, and open programming standards. Mobile phones are heterogeneous, employing varying input methods, display screens, browsers, operating systems, and user interfaces. Mobile devices do not conform to any standards.
- Mobile search usability issues include smaller screens, typing limitations of phone keypads and the cost of spending time scrolling through mobile search results.
- Mobile devices are currently less likely to be used for general browsing but more for retrieving specific information.

These limitations describe the problems that need to be taken into account when developing an acceptable mobile search and advertising procedure. In addition to these immediate differences, future mobile search applications may be able to capitalize on user specific information [7]. Mobile phones are increasingly associated with each individual’s personal style, representative of their owner’s personality, with specific demographics, behaviour patterns, and personal interests etc. This information offers the opportunity for more relevant search results to be determined. Mobile phones may also be able to leverage location as an additional search parameter, allowing for greater specificity for search results.

**Reading this paper will tell you the following:**

- The difference between mobile web search and PCs web search.
- Why mobile dialog is needed?
- What is mobile advertising?
- How mobile search and mobile advertising will work together.

---

**QAMEN versus Mobile Web Search**

---

The mobile web consists of web pages that are designed specifically for display on mobile devices. Due to their limited capabilities (relative to standard computers) mobile devices access and render web content using the specialized **Wireless Application Protocol (WAP)**. WAP is separate and distinct from the **Hyper-Text Transfer Protocol (HTTP)** that computers use to access HTML pages. As such, WAP browsers can only access pages that are written in **xHTML (eXtensible Hyper-Text Markup Language)** or **WML (Wireless Markup Language)**. WAP 1.x

browsers access only WML pages, while WAP 2.0 browsers access both XHTML pages and WML pages. The mobile web is determined by the universe of content that is written in WML or XHTML.

Searching the mobile web returns only WAP pages. The world is short of quality WAP sites in general. Most WAP sites are poorly designed, under-used, unstable and limited in content and service. There are WAP search engines that index the mobile web but many do not work well due to the volatile nature of the current mobile web – there are many outdated and redundant links. Furthermore, most WAP site crawling technologies are underdeveloped [10].

QAMEN is an HTML compatible search engine. It works almost the same as computer browsers and can access HTML search pages from the huge internet webpage databases in the same way that computer browsers do. In providing the result of internet searching QAMEN converts web pages to a simplified format compatible with handheld devices. Search technologies of QAMEN are evolving to provide users with appropriate results despite the often unstructured web content.

### QAMEN's Basic Commands

By default any request is considered by QAMEN as a request for searching in the local knowledge base, and/or in the Internet. Mobile question answering differs from standard information retrieval methods in some important respect. Firstly, it needs to retrieve specific fact information rather than whole documents. Secondly, it should select among the found facts the shortest and appropriate fact to meet the requirement of 160 characters. In short what a user normally requires is a precise answer to a question. But the concept "precise answer" is very fuzzy because of its reflects potential ambiguities in a user's request. For example, QAMEN even theoretically cannot define what is the precise answer to a user's request: **"I'm looking for Hilton address in London"** (see Fig. 1).

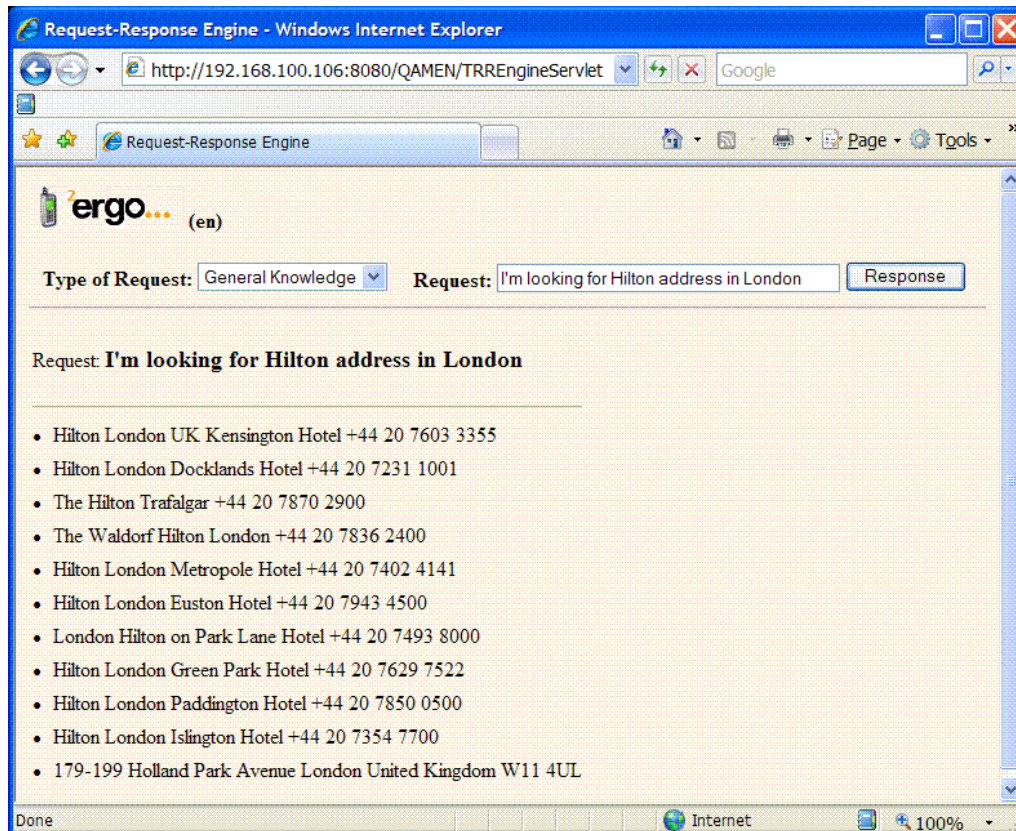


Figure 1. List of locations and phone numbers of Hilton hotels in London

On the one hand, the most important, the result must be relevant to the user's search, on the other hand, there is no guarantee that the displayed response: **"The Hilton Trafalgar +44 20 7870 2900"** exactly meets the user's expectation. Enhancing the user profile, of course, could significantly improve the selection of appropriate answer. Even an unexpected advert along with the correct response might improve user satisfaction.

The mobile search industry is still in its infancy and the primary barriers, cited by subscribers as affecting potential uptake, are not knowing how to use the search engine (23%) and not thinking about using mobile search on their phone (19%). Just 20% of UK subscribers actually search content on the mobile internet, despite an industry perception that 89% do [11]. This gap between reality and perception suggests that the industry needs to promote the benefits of mobile search more actively as well as educating consumers on how mobile search may be used to find relevant content.

Taking into account such a situation QAMEN provides users with the maximum flexibility for selection of the desirable combination of options. Mobile users seek quick and convenient access to information and services. The following basic commands, for example, allow a user to select any combination of options online:

- Type in **Advert ON** to activate advertising (or **Advert OFF** to quit).
- Type in **Dialog ON** to activate dialog (or **Dialog OFF** to quit).
- Type in **Person ON** to activate personalization (or **Person OFF** to quit).
- Type in **Language BG** to allow user to use Bulgarian language (or **Language EN** to back to English).
- Type in **Election ON** to select an application domain with data regarding the UK General Election 1997, 2001 and 2005 (or **Election OFF** to quit).

For example, in the result of Dialog activation when QAMEN provides interactions between itself and the user the response to the same request: **"I'm looking for Hilton address in London"** will be represented differently (see Figure 2). Space on the mobile phone screen is at a premium, and users have limited input mechanisms, so any result of mobile searching needs to be easy to navigate using only the mobile phone's own keypad.

Each user can have multiple active mobile dialogs. QAMEN holds continually the personal profile of each user as well as the current state of dialogue for each search activity.

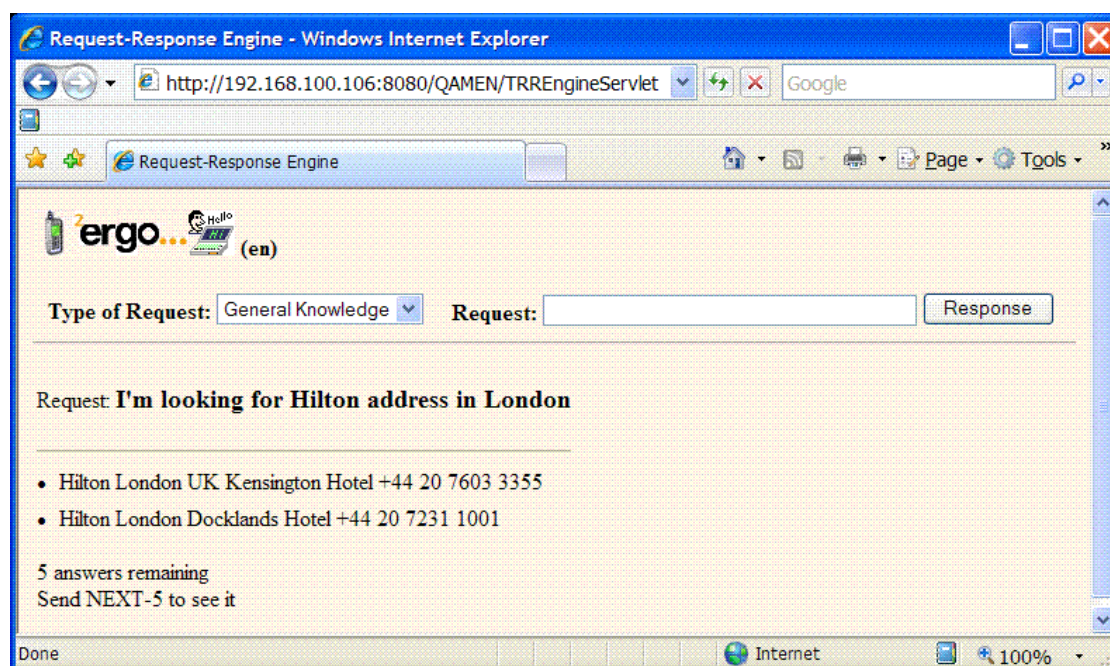


Figure 2. List of locations and phone numbers of Hilton hotels in London with dialog

Let us distinguish **Answers Quantity (AQ)** and **SMS-Responses Quantity (RQ)**. For considered request  $AQ=11$  (see Figure 1). RQ represents number of SMS, which need to be sent to cover AQ. As a rule the length of one response less the size of a SMS. That is why one SMS may include several answers i.e.  $RQ \leq AQ$ . QAMEN shows how many RQ left (see Figure 2).

---

## Mobile Advertising

---

Mobile advertising can basically be defined as the business that using the wireless channel as a medium to deliver the advertisement message or slogan encourages people to buy products and services. Mobile advertising is not simply linking users to websites by clicking on banners or pop-up windows through their mobile phone browser. It offers the ability to go beyond the mainstream of electronic advertising, which mainly targets the mass market predominantly over the Internet, to allow businesses to deliver their particular messages to specific individuals at specific locations.

QAMEN is an ideal way for advertisers to reach target markets and establish a one-to-one relationship with the consumer, a significant objective for all advertisers. For example, weather application is defined as personalized, localized weather prediction according to user location and personal profile. Weather related advertisement system knows how to match the right advertisement to the right weather in the most effective manner. For example, implementing a decision to start a soft drink campaign when the temperature approaches,  $32^{\circ}\text{C} / 90^{\circ}\text{F}$  according to user location (if the user is close to the beach and experiencing higher levels of effective temp he will experience a different ad in different temperatures). Such approaches would help the advertiser to optimise its advertising campaign.

The following key characteristics drive the success of messaging as a source of mobile advertising inventory [12]:

- **Ubiquitous SMS access** – Virtually all mobile phones can receive SMS, and the majority of users use SMS on a regular basis. Today, SMS is the most widely used mobile phone service after voice.
- **High attention level** – Users almost never delete mobile message without opening them and reading at least part of content.
- **Engaging** – Once displayed, an effective advert can engage users directly in various ways, such as interactive message reply, click to transfer to the Web, or click to call.
- **Compatibility** – Messaging usually works between different networks and between different countries.
- **Response collection is easier** – It also may be achieved immediately. Brands may have access to real time response information and may modify the campaign according to the results, long before a campaign terminates.
- **Direct and personal way of communication** – Customer has the sense of feeling that the advert addresses only him/her.

---

## Mobile Advertisement Creation

---

In this paper we consider adverts that are text only. Mobile messaging represents an opportunity for advertising placement. Advertisement can be inserted in SMS content that subscribers request and receive by using the free (non-used) space, up to the message size limit. The main task of advert creation is to establish an advert structure. There are, as yet, no standards for representing these, because there is no general agreement on what an advert should contain. That is why we were free to offer our vision of advert structure. First, let us describe the general requirements, restrictions and conditions to advert creation:

- A poor mobile advertising conveys a bad impression of any company, potentially turning away customers.
- An advert must be related to the original message content i.e. text adverts have to be shown in response to keywords entered on user queries.
- Mobile advertising is moving more towards dialogue type of advertising. Limiting user input to numeric or short sequence of text due to the limitation of the keypad on most mobile phones.
- Adverts should provide click-to-call links that allows the users to make a phone call directly from the displayed advert.
- Analysis of advert “activities” is a crucial feature in their future development in order to automate an improved process of advertising. The correctly specified metrics for an advert should include:
  - Total adverts sent.
  - Total adverts delivered.
  - Unique deliveries (distinct mobile phones).
  - Content of advert served.
  - Type of advert (independent or request dependent).
  - Location of served advert.
  - Date and Time when advert has been displayed.
  - Season of advert displaying.
- SMS advertising is appended to the bottom of the content message to be sent to the user. The available space for each advert is therefore dependent on how much space is left after subtracting the characters used in the main content body of the message.
- To optimize an advert placement:
  - The advertiser should develop several versions of adverts to be used, with the optimum length directly dependent on the length of sender’s message.
  - Use punctuation only if required for clarity or emphasis.
  - Try to avoid a carriage return because it may count as two characters.

Let us distinguish between two types of advert: 1) **independent** if there is no any link with areas, season or location, and 2) **dependent** otherwise. **For both types of adverts link with user request is required.** Advert can be easily created and updated via the web-based interface (see Figure 3).

The advert structure consists of the following items:

- **Areas:** *game, music, video, product, weather, sport, health, finance, holiday, transport, omit.* For advert several areas might be selected e.g. *sport* and *weather*.
- **Seasons:** *all seasons, winter, spring, summer, autumn, omit.* If *spring* is selected the months *March, April* and *May* will be added to advert pattern automatically.
- **Location:** *country, county, city, omit.*
- **Slogan** is a short (< 100 characters) memorable text advert.
- **Advert’s Key Words (AKW)** include not only synonyms of slogan’s keywords but also describes the conditions for use of the advertised product. For example, if “ice cream” is advertised the words: *hot* and *heat* should be added to keywords.
- **Advertiser** and its **contacts:** *country, city, address, web site, email, telephone.*



- **Priority** depends on payment and its value lie within 1-10 range.
- **Quantity** shows how many times advertiser wants advert to be displayed and (in brackets) how many times advert has been displayed.

Before storing the advert description in the **Adverts DataBase (ADB)** which is the main part of advert – (the **Advert Pattern (AP)**) – is created. The general form of AP is the following:

AP=<SKW+AKW><Areas><Seasons><Location><Priority><Quantity>, where:

- **SKW means Slogan's Key Words**. SKW are extracted from the Slogan and together with AKW are saved in the advert dictionary;
- Meaningful selection (i.e. when the selected value is no equal *omit*) of **Areas, Seasons** and/or **Location** are saved in the advert dictionary as well. The selection of *All Seasons* means that all twelve months and four season names will be included in the dictionary.

Figure 3. Web-based interface for Advert creation and editing

It is important to emphasise that any words are inserted in the advert dictionary just once but that the link chain related to any word will be extended, and will consist of the information of all adverts from ADB where this word has been involved. For example, word *May* has a link to the set of adverts *Seasons* that have been described as

*All Seasons, or Spring.* An example of advert, which is extracted from the ADB in accordance with user's request is shown in Figure 4.

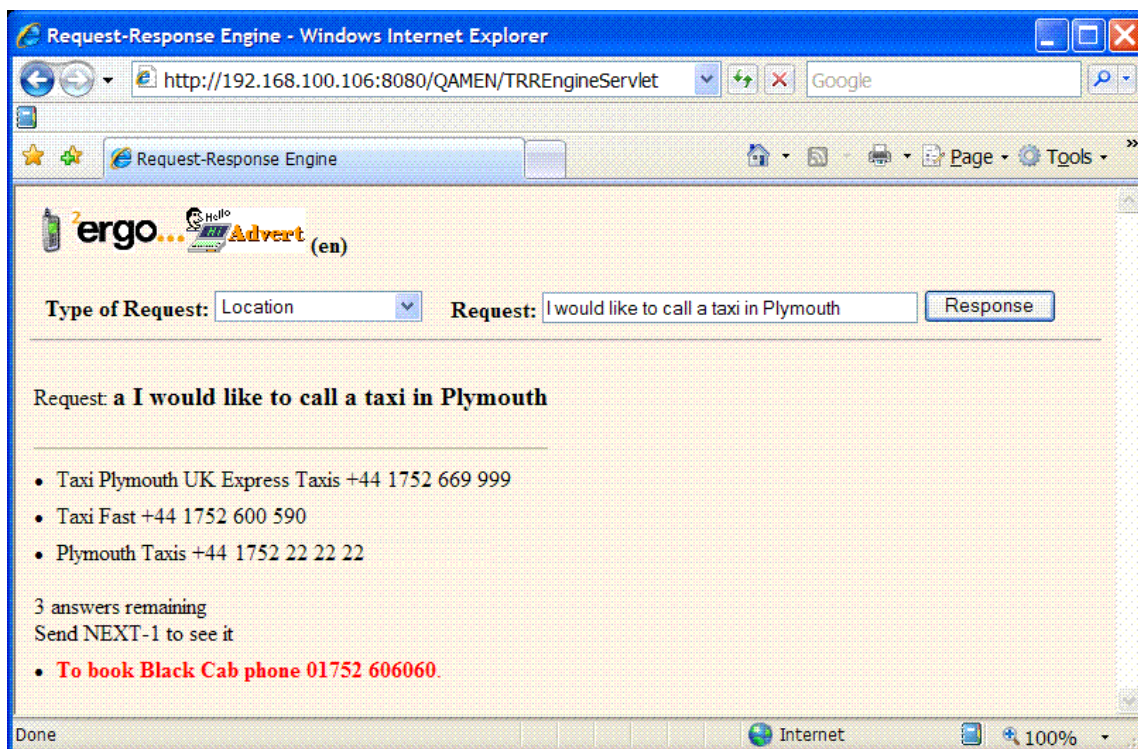


Figure 4. Example of Advert

### Algorithm of Advert Selection

Let us consider the algorithm of advert selection:

1. QAMEN should be ready for advert i.e. command **Advert ON** should be sent from the mobile phone.
2. User type in and send the request e.g. *I would like to call a taxi in Plymouth* (see Figure 4).
3. QAMEN extracts KW from the request (RKW) i.e. *call, taxi and Plymouth*.
4. Using advert dictionary QAMEN extracts from ADB the list of advert patterns and calculates their "weights" i.e. APW. For APW calculation it is just enough to find intersection RKW with AP from ADB. If RKW belongs to SKW then 2 points is added to APW. In all other cases i.e. when RKW represents **Areas, Seasons, Location** explicitly, or belong to AKW just 1 point is added to APW. For the request in the example APW=3 but for request: *"I'm looking for phone No to book a taxi in Plymouth"* APW=6.
5. QAMEN makes descending sort of AP in accordance with their APW. If two AP have the same APW the **Priority** of AP defines the order. If **Priority** is equal the **Quantity** of AP and **Date** and **Time** when adverts have been displayed the last time define their order.
6. QAMEN calculates the free space for advert and in the result of calculation select an appropriate version of advert (see Figure 4).
7. If Web searching produces no response then just the advert can be displayed (see Figure 5).



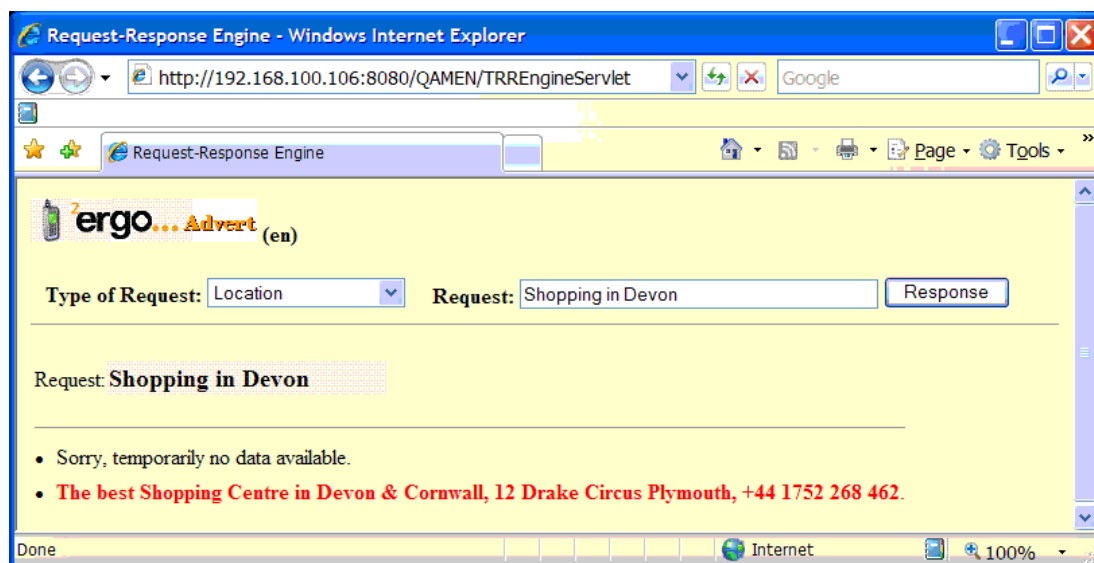


Figure 5. Web searching was not a success

## Conclusion

We believe that mobile search and mobile advertising offers a significant opportunity for many players in the wireless industry. While consumers across the world use Internet search applications as their primary access point to information and web sites, mobile users are just starting to experiment with the potential of mobile search and advertising applications. In this paper we describe our vision of mobile search and advertising to improve the efficiency of SMS. We turned our particular attention towards mobile advertising and its possible application in a mobile environment. The object of our research is to improve query response by adding adverts. It is important to offer and realize some ideas (not necessarily the best) when there are as yet no standards for representing mobile adverts. Of course, the ultimate criterion of "good" mobile search and advertising is that a user should be satisfied with search results and subscribe to the content of adverts. Future studies into this aspect of mobile phone use could be conducted as a series of further case studies in other countries in order to facilitate comparison: we have already started to create the Cyrillic version of QAMEN. The first step has been made in the creation of a Bulgarian version of QAMEN (see Figure 6).

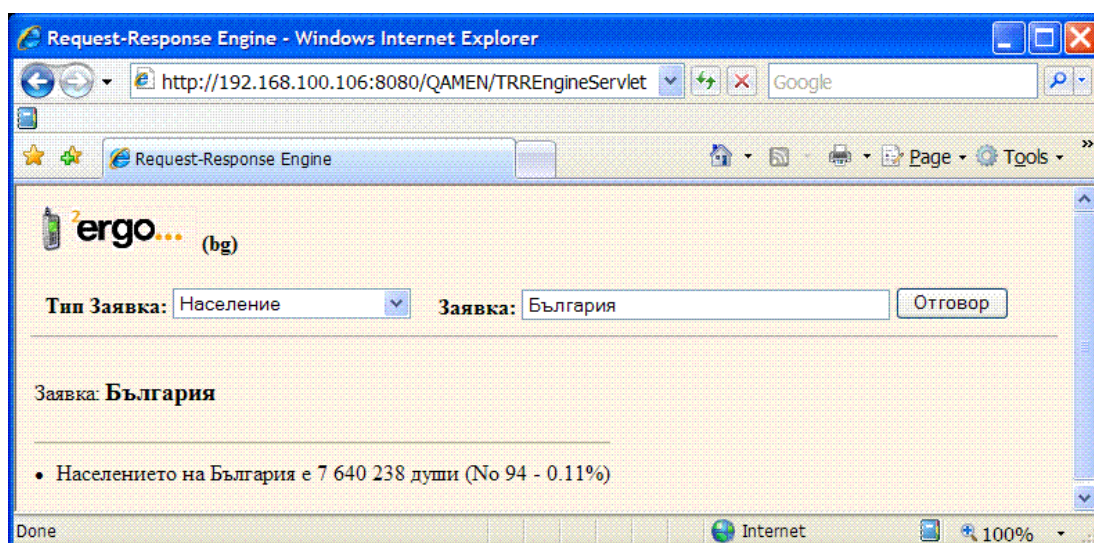


Figure 6. Bulgarian version of QAMEN

---

## Bibliography

---

- [1] G.Coles, T.Coles, V.A.Lovitskii, "Natural Interface Language", *Proc. of the VIII-th International Conference on Knowledge-Dialogue-Solution: KDS-99*, Kacivelli (Ukraine), 104 -109, 1999.
- [2] T.Coles, V.A.Lovitskii, "Text Searching and Mining", *Journal of Artificial Intelligence, National Academy of Sciences of Ukraine*, Vol 3, 488-496, 2000.
- [3] D.Burns, R.Fallon, P.Lewis, V.Lovitskii, S.Owen, "Verbal Dialogue Versus Written Dialogue", *International Journal "Information Theories & Applications"*, Vol 12(4), 369-377, 2005.
- [4] Ken Braithwaite, Mark Lishman, Vladimir Lovitskii, David Traynor, "Distinctive Features of Mobile Messages Processing", *International Journal "Information Theories & Applications"*, Vol 14(2), 154-160, 2007.
- [5] Guy Francis, Mark Lishman, Vladimir Lovitskii, Michael Thrasher, David Traynor, "Instantaneous Database Access", *International Journal "Information Theories & Applications"*, Vol 14(2), 161-168, 2007.
- [6] Vladimir Lovitskii, Michael Thrasher, David Traynor, "Automated Response To Query System", *International Journal "Information Theories & Applications"*, Vol 15(2), 143-152, 2008.
- [7] Lee Johnston, Vladimir Lovitskii, Ian Price, Michael Thrasher, David Traynor, "Personalized Question-Answering Mobile System", *International Book Series: "Information Science and Computing"*, book 2: "Advanced Research in Artificial Intelligence", 123-132, 2008.
- [8] Loren Baker, "Mobile Search Advertising", *Search Engine Journal*, February 2008  
<http://www.searchenginejournal.com/mobile-search-advertising-to-hit-38-billion-by-2012-14-billion-in-us/6382/>
- [9] <http://www.mmaglobal.com/uploads/MMAMobileSearchIntro.pdf>.
- [10] "Mobile Search and its Implications for Search Engine Marketing" provided by <http://www.oneupweb.com>
- [11] Eden Zoller, Ovum, <http://www.ovum.com/go/content/c,432,35105>
- [12] <http://www.iabmexico.com/downloads/mobileadoverview.pdf>.

---

## Authors information

---

**Vladimir Lovitskii** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [vladimir.lovitskii@fsmail.net](mailto:vladimir.lovitskii@fsmail.net)

**Colin McCaffery** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [colin.mccaffery@2ergo.com](mailto:colin.mccaffery@2ergo.com)

**Michael Thrasher** – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, e-mail: [mthrasher@plymouth.ac.uk](mailto:mthrasher@plymouth.ac.uk)

**David Traynor** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [david.traynor@2ergo.com](mailto:david.traynor@2ergo.com)

**Peter Wright** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [peter.wright@2ergo.com](mailto:peter.wright@2ergo.com)

---

## MOBILE ELECTION

**Elena Long, Vladimir Lovitskii, Michael Thrasher, David Traynor**

**Abstract:** *Mobile phones have the potential of fostering political mobilisation. There is a significant political power in mobile technology. Like the Internet, mobile phones facilitate communication and rapid access to information. Compared to the Internet, however, mobile phone diffusion has reached a larger proportion of the population in most countries, and thus the impact of this new medium is conceivably greater. There are now more mobile phones in the UK than there are people (averaging at 121 mobile phones for every 100 people). In this paper, the attempt to use modern mobile technology to handle the General Election, is discussed. The pre-election advertising, election day issues, including the election news and results as they come in, and answering questions via text message regarding the results of current and/or previous general elections are considered.*

**Keywords:** *mobile text messages, mobile election, mobile advertising, question-answering system*

**ACM Classification Keywords:** *1.2 Artificial intelligence: 1.2.7 Natural Language Processing: Text analysis.*

**Conference:** *The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009*

---

### Introduction

In our previous papers [1,2] the Question-Answering Mobile ENgine (QAMEN) has been represented. QAMEN is based on industry-standard SMS messaging technology and thus works with any mobile in any GSM network. In this paper QAMEN will be restricted by General Elections issues. For clarity we will distinguish two versions of QAMEN for election: PC version – QAMEN-E<sub>PC</sub> and Mobile version QAMEN-E<sub>MBL</sub> and only the **Election** Application Domain (AD) will be considered.

While it is too early to determine the political effects of mobile phone diffusion, the political events in different countries suggest that mobile technology may come to play an important role in political participation and democracy. Text messaging has already played an important role in the 2008 United States campaign of Barack Obama. Obama's text message announcing his vice-presidential candidate selection of Joe Biden reached approximately 2.9 million Mobile Subscribers (MS), many of whom signed-up with the promise of receiving vital information via text messaging on their mobile phone.

We are definitely against blind copying the digital techniques used in the 2008 US presidential election. One disincentive is the short campaigning time of UK general elections. US presidential campaigns are two years long and so the candidates were able to test different digital media. But with UK general election campaigns only three and a half weeks long, UK parties will have to get it right, first time. Planners argue that use of digital media will come down to the demographics of swing voters. It depends on whether the parties believe that the next election will be won by retaining the loyalty of the over-50s or at the margin of first-time and under-25 voters, most of whom are mobile-literate.

Despite the differences, the growth of the importance of technology in politics cannot be ignored. UK politicians are still learning the new tone, style and dynamics needed for mobile campaigning. Politicians need to see it as an opportunity to express themselves, rather than a threat; i.e. politicians need to overcome their fear of the unfamiliar and embrace the opportunities it provides.

The Electoral Commission has warned that the UK electoral system continues to operate under 19th century structures and requires urgent reform before the next general election,. "The planning and running of elections need to be more robust and coordinated," said Sam Younger, chairman of the Electoral Commission. "We are still trying to run 21st century elections with 19th century structures, and the system is under severe strain." [3].

In the UK, legislation currently prevents electronic voting in general elections, but the technology has been well tested in 17 pilot projects during local and European elections. These trials were funded by the Office of the Deputy Prime Minister, which is also investing £12m into Core (co-ordinated online register of electors), a project that aims to modernise electoral rolls - a crucial step in enabling national e-voting systems.

The biggest e-voting trial conducted so far in the UK was in Sheffield, where 174,000 citizens were given the opportunity to vote using the internet, kiosks and mobile phones. Voters in the city were given the choice of voting using a traditional ballot paper, a mobile phone text message, a touch-tone telephone, a website or a touch-screen internet kiosk at a polling station.

The Electoral Commission wants e-voting technology to be proven before it will give the green light to an electronic general election, but proving the technology is difficult without a large-scale trial. The Office of the Deputy Prime Minister says there will be no large-scale e-voting projects until "issues of secrecy, security and technological penetration have been addressed". This makes it into a **vicious circle**. In this paper the way of breaking such a circle will be discussed. QAMEN-E<sub>MBL</sub> will be represented as an election monitoring system. The purpose of this paper is to describe the main stages of mobile election and discuss the ways of mobile election monitoring.

**Reading this paper will tell you the following:**

- Mobile election overview.
- Pre-election advertising.
- The Election Day.
- Mobile Election Results.
- QAMEN-E<sub>MBL</sub> Versus QAMEN-E<sub>PC</sub>.
- Mobile Request Processing.

---

**Mobile Election Overview**

---

- The first local and mayoral electoral vote in the UK by text message took place on 23rd May 2002 [4].
- Korean campaigners sent text messages to encourage voting in 2002, while the Italian government in 2004 sent a reminder message about upcoming European elections to all mobile phone subscribers [5].
- Mr. Tony Blair became the first UK Prime Minister to use text message technology to talk directly to the people on 25th November 2004 [4], answering questions submitted in advance by text message from members of the public as well as in real-time in a mobile phone chat-room, transmitted live from No.10 Downing Street.
- The Spanish general election of 2004 occurred in the wake of an unprecedented terrorist attack, but its outcome reflects the potential that mobile phones have to provide the user with independent information and bring about voter mobilisation. Mobile technology contributed to the quality of democratic practices in the sense that SMS messages helped provide citizens with more information about the rapidly unfolding events related to the terrorist attacks, including the reaction by the government and the opposition party as well as the investigation and the protests during the "day of reflection" [6].
- The information about the first mobile phone election in the UK had been announced in April 2005 [7]. In the UK, the General Election always sparked a frenzied dash for information as voters tried to keep up to date with hot election issues. As with all major news stories, people want to be able to follow events as they unfold and the mobile phone is proving to be the ideal tool for this.
- On May 21, 2006 the Montenegro, a small country in the former Yugoslavia, saw the first instance of volunteer monitors using SMS, as their main election reporting tool. A Montenegrin NGO (Non-Governmental Organization) was the first organization in the world to use text messaging to meet all election day reporting requirements.

- Since then, mobile phones have been deployed in six elections in countries around the world, with volunteers systematically using text messaging in election monitoring. SMS monitoring is becoming a highly sophisticated rapid-report tool, used not just in a referendum election like in Montenegro, but in parliamentary elections with a plethora of candidates and parties and complex data reported via SMS. This was the case in Bahrain, a small country in the Middle East, where monitors reported individual election tallies in a series of five to forty concurrent SMS messages, using a sophisticated coding system, with near accuracy.
- In the two years since the first large-scale SMS monitoring in Montenegro, there have been rapid improvements in mobile services as competition in the wireless industry has increased worldwide, and there is growing interest and understanding on the part of NGOs that systematic election monitoring is not as difficult as it first may seem. As election monitoring via SMS becomes standardized and NGOs gain experience, there is no reason for mobile phones and SMS not to play a greater role in other areas of civic participation.
- Election in Sierra Leone [8] is lead by the National Election Watch (NEW), a coalition of over 200 NGOs in the country. NEW has monitors at 500 of the 6171 polling stations. Monitors report on whether there are any irregularities via SMS back to headquarters.
- The Estonian Government has passed a new bill that will allow its citizens to vote using their mobile phones in the next election [9]. The measure will come into effect for the 2011 election, and makes Estonia the first country in the world to approve such a method. In order to vote using their mobile phone, Estonians will be required to purchase a special chip for their handsets which will verify the voter's identity and authorise their vote on the system.

The future is bright for innovative ways in which mobile phones are used by citizens to participate and engage in their countries as the mobile revolution unfolds. In fact, the 2008 US presidential election saw the widest possible mix of offline and online media used to help candidates connect with some 200 million voters scattered across America's vast expanse. Experts are predicting that in future elections, the use of mobile technology will become the standard and play an even larger role than in recent US election where Barack Obama's use of text messaging is already being held as a successful way to reach out to the constituents.

Mobile campaigning needs a new political vocabulary, style and humour - far removed from the stultifying prose of traditional party literature. Mobile is all about personalized content; offering politicians a way of communicating with voters that is simply not possible with television or newspapers.

---

### **Pre-election Advertising**

---

Election advertising is defined as any content that can reasonably be regarded as attempting to gain electoral success for any candidate or political party which seeks to increase their status or position. Election advertising is anything that can reasonably be regarded as influencing a voter in a particular direction, and it's a media-neutral definition. Election advertising requires advertiser identification, and spending limits apply. A message supporting someone else needs the beneficiary's written permission. The Election Advertising Campaign (EAC) will provide a unique perspective for both candidate or political party, and MS. EACs goal is to get more votes, more attention and more exposure for the General Election 2010 in the UK.

What can the General Election campaign, political parties, candidates and Registered Mobile Subscribers (RMS) expect from EAC?

- Increased voter turnout due to text message reminders.
- Votes by text could increase turnout, especially a potential record number of young voters. 95% of 16-24 year olds use text messaging regularly, each sending an average of 400 texts per month. Young Britons would be far more willing to vote in the General Elections if the government were to introduce voting by mobile phone. ICM Research found a huge 73% of 18 to 24-year-olds would have voted if they had the option of doing so by mobile phone. The opportunity of being able to cast a vote via text message would make people more likely to participate in an election.

- Political parties, for instance, might invite party members and ordinary citizens to participate in the pre-election candidate nomination process by mobile phone voting.
- Candidates will be able to use text messages to collect campaign contributions. Mobile search is the ideal opportunity for candidates to position themselves about key issues like jobs, schools, knife crime, etc.
- RMS will be able to receive text messages from their candidate or party of choice. This is another sign of how political outreach could try to keep up with the changing nature of personal communication.
- Every evening at 8pm a text summary of the key political events of the day and the latest election news will be sent to the RMS.
- Receiving a reminder for scheduled and special elections events could become the norm.
- Be the first to know with breaking news alerts.
- EAC keeps RMS abreast of the fast changing political scene in the final days before the General Election.
- RMS might be asked from time to time for whom they are likely to vote, to define the current situation of candidates popularity as a projection of the election results.

Registration as RMS is very simple:

- Type in your name, post code and password.
- Send this message to 5-digit short code. It should be memorable short code like ELECT (35328), or 2ergo (23746). It is important to note that these short codes should only be considered as examples.
- Confirmation about registration will be sent as a text message to the RMS mobile phone along with a Pin number (required for election day only) and notes about what kind of information will be regularly sent to the RMS mobile phone. All information, of course, needs to be suitable for the limited text messaging format of 160 characters.
- Simply send a text message CANCEL to 35328, or 23746 and you will be unsubscribed from all services.

For non-registered MS it would be enough to text NEWS to 35328, or 23746 for breaking news alerts, or text EVENTS to the same short codes for the nearest election events, or text RESULT for actual election results.

In order for EAC to succeed, a number of conditions have to be in place:

- The power of the message lies beyond 160 characters: One of the best tactics of wise political campaign managers has been the use of embedded links in each message that direct to a candidate's homepage.
- Political campaigns have utilized demographic information to target voters in particular cities, regions or with specific issues in mind. With location-based services becoming more commonplace on phones, the ability to deliver targeted messages will be one of the biggest enhancements to mobile messaging in the near future.
- QAMEN-E<sub>MBL</sub> should be able to handle thousands of messages per minute to one mobile short code.
- The mobile provider (e.g. 2ergo for QAMEN-E<sub>MBL</sub>) would give the highest priority to the election mobile short code.
- Note that each SMS message sent to QAMEN-E<sub>MBL</sub> for this service will cost 10-15p (network charges vary).

An important consideration is the cost of a wide-scale program. An analysis of the last years Presidential Election in the United States,[10] allows the surmise that not only has text messaging surpassed actual calls as the most popular form of mobile communication but its cost-effectiveness in elections was astounding. For a political campaign, the acquisition cost of an "opt in" text user (a RMS in our case) is about 13 cents. By comparison, the acquisition cost of door-to-door canvassing is about \$2.50 and for phone calling, it's about \$1.00. The difference in actual cost-per-vote results is even more obvious: \$1.62 per vote for opt-in text lists compared to \$20-\$35 per vote from phone calls, leaflets and door-to-door visits.



However, the Sheffield trial did reveal some potential problems with e-voting. The biggest problem was cost. Offering voters access to secure and robust electronic voting channels more than doubled the cost of running the election, to £55 per voter, according to the Institute of Public Policy Research.

Mobile phone technology allows users to be directly accessible at all times and locations and that is why it would be more effective for political parties to use the power of mobile messaging, namely QAMEN-EMBL, not only during the short campaigning time of UK general elections but the whole year.

### The Election Day

RMS from around the country used the power of text messaging to cast their vote. Attendees cast their ballots by texting only the name of the candidate to the short code 2VOTE (28683), or ELECT, or 2ergo to take part in the election.

A note about security. RMS voting requires the entry of a password and pin. In a traditional voting system no formal identification document is required when voters present themselves at a polling station; and yet the same person may be required to present three forms of ID just to register at their local Blockbuster video store. Therefore QAMEN is a vast step forward in RMS security.

Arriving messages are passed to QAMEN-EMBL and then the data is compiled in a database (DB) ready for analysis. It is amazing to see the numbers change on the screen as the SMS messages pour into the DB (see Figure 1). The RMS will be kept updated on the latest election news, latest announcements and sent reports on quantitative data such as real-time voter turnout and even on the actual election results. For non-registered MS it would be enough to text RESULT to 35328, or 23746 to get the latest election results anytime and anywhere.

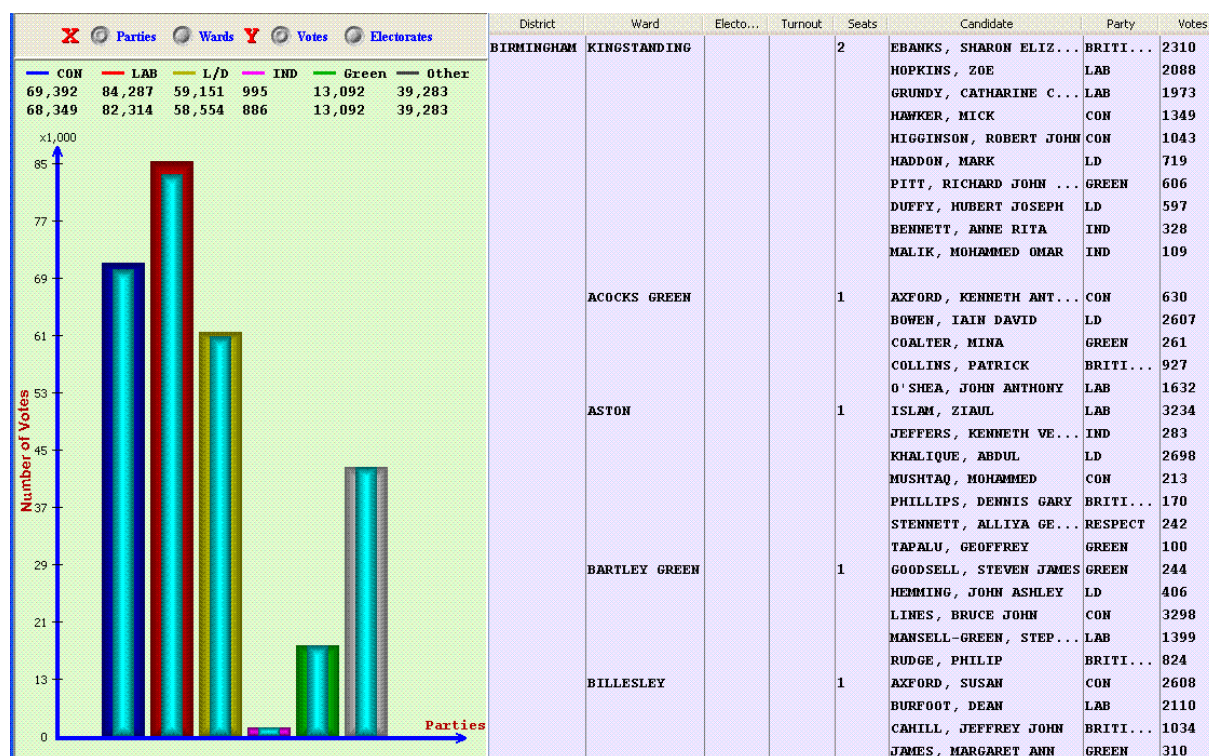


Figure 1. Current result of local election in Birmingham

### Mobile Election Results

QAMEN-EMBL is supposed to make nationwide election results available not only to RMS but to all mobile devices. MS can send any request regarding election results to 35328, or 23746 and QAMEN-EMBL will reply via text message.

The visualization of election results can make it more intuitive and productive. People have always relied on visual tools such as maps, charts, and diagrams to better understand problems and solve them quicker. The mobility context and technical limitations such as a small screen size make it impossible to simply port visualization applications from desktop computers to mobile devices, but researchers are starting to address these challenges. Considerable effort is needed to understand how to design effective visualizations for mobile devices. Although many researchers have proposed specific techniques, no reports on the topic have yet provided a broad discussion of mobile visualization that could be useful to mobile application developers. In this paper only textual presentation of election results will be considered.

Mobile phones (and consequently QAMEN-EMBL) have some limitations when compared to PCs (QAMEN-EPC). These limitations are the problems that need to be taken into account when developing an acceptable mobile question-answering procedure.

### QAMEN-EMBL Versus QAMEN-EPC

Compared to QAMEN-EPC, QAMEN-EMBL has many restrictions that have to be considered and overcome:

- Displays are very limited due to smaller screen size, the 160 characters SMS restriction and lower resolution.
- Input peripherals such as tiny keypads, micro joysticks, and rollers are often inadequate for complex tasks.
- Connectivity is slower, affecting interactivity when a significant quantity of data is stored on remote databases.
- QAMEN-EPC might provide the powerful command prompt (see Figure 2).
- It is easy to use such a prompt to represent requests in clear, grammatical and correctly-spelt language. Often Mobile Request (MR) for QAMEN-EMBL will be ungrammatical. As a rule MSs:
  - Do not want to use upper case to type request like “*george bush, washington dc*” [11]. Or use dots to separate “*d*” and “*c*”.
  - Dropping ‘?’ at the end of MR.
  - Not using any punctuation at all.
  - Deletion of articles.
- The fact is that MR simply will not be spelt, punctuated, and capitalised correctly but the main requirement for QAMEN-EMBL is - to handle non-standard or poorly formed/structured (but, nevertheless, meaningful) user's MR.
- QAMEN-EPC has no problems displaying the response in a convenient format for the user(see Figure 3) but for QAMEN-EMBL the small screen and 160 characters restrictions may cause some problems when it comes to displaying the response.

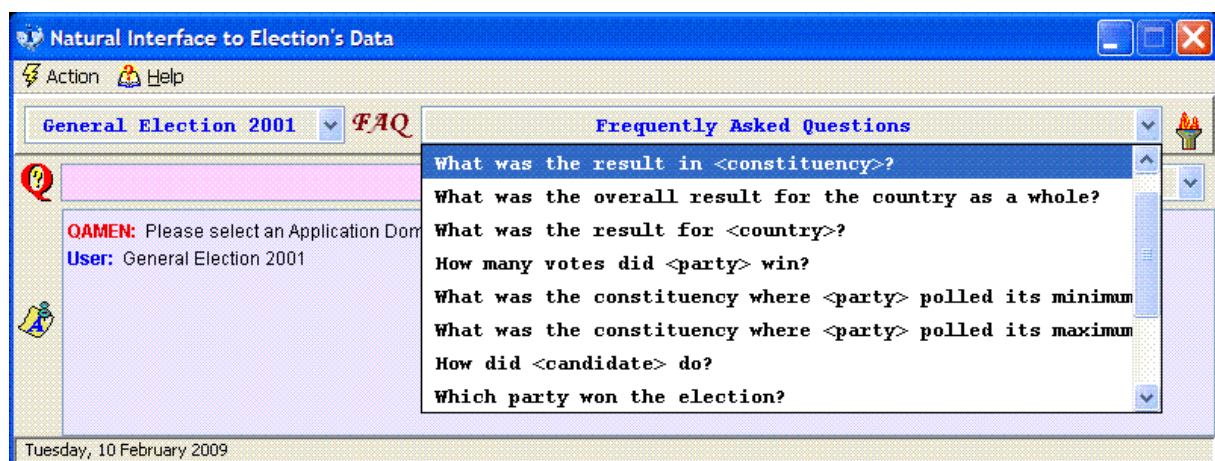


Figure 2. QAMEN-EPC: Command prompt interface



The main conclusion from such comparison is QAMEN-E<sub>MBL</sub> should be **more intelligent** than QAMEN-E<sub>PC</sub>. More evidence for such conclusion comes from the fact that it is simply impossible to require the users to remember, for example, the exact name of the constituency in order to correctly ask what seems a very simple question: "Who won the election in Suffolk Central & Ipswich North in 2001?". It would be expected that the user instead of using the symbol '&' types in "**and**". In which case QAMEN-E<sub>PC</sub> will not find the constituency in DB and will have to generate the clarification dialog:

**QAMEN:** Do you mean Suffolk Coastal, Suffolk South, or Suffolk West constituency?

**User:** No, I mean Suffolk Central.

**QAMEN:** Suffolk Central constituency does not exist but there is Suffolk Central & Ipswich North constituency.

**User:** It's exactly what I meant.

**QAMEN:** Thank you.

Theoretically QAMEN-E<sub>MBL</sub> can do the same, but such dialog would not be acceptable due to time and money. But QAMEN-E<sub>MBL</sub> is an intelligent system and in the result of similarity measurement [12] between MR *Suffolk Central and Ipswich North* and similar DB values namely: *Suffolk Central & Ipswich North*, *Suffolk Coastal*, *Suffolk South*, *Suffolk West* and *Ipswich* constituencies, QAMEN-E<sub>MBL</sub> selects *Suffolk Central & Ipswich North* with great confidence.

In the case when user simply made a mistake and instead of typing in the desirable constituency *Hereford* in the MR: "Who won the election in Hereford" he/she entered *Hertford* (it's **wrong** but at the same time it's **right** from the QAMEN-E<sub>MBL</sub> point of view because it has the right part of an existing constituency), QAMEN-E<sub>MBL</sub> found an answer for the constituency *Hertford & Stortford*. When MS sees the response, he/she realises that MR was wrong and corrects it.

**General Election 2001** **FAQ** **Frequently Asked Questions**

What was the result in <constituency>? **New** **Constituencies**

**QAMEN:** Please select an Application Domain.  
**User:** General Election 2001  
**User:** What was the result in Plymouth devonport?

**General Election 2001**

Constituency:	Candidates	Votes
<b>Plymouth devonport</b>		
<b>Parties</b>		
Lab	Jamieson, David	24,322
Con	Glen, John	11,289
LD	Baldry, Keith	4,513
UK Ind	Parker, Michael	958
Soc All	Staunton, Tony	334
Soc Lab	Hawkins, Rob	303

**QAMEN:** Any more questions?

Tuesday, 10 February 2009

Figure 3. QAMEN-E<sub>PC</sub>: Question-Answering interface

## Mobile Request Processing

The main purpose of MR processing is to understand **what was meant** rather than **what was said** and in the result of such, understanding and returning to MS only elections-relevant information. The mechanism of query phrasing is very simple: "eliminating the unnecessary until only the necessary remains" and has been discussed elsewhere [13]. Here we just remind ourselves of the main steps involved in MR processing.

- QAMEN-EMBL takes the MR as a character sequence and converts the original MR to a *skeleton* by noisy (non-searchable) words elimination. In the result of such conversion MR will contain only **meaningful** words. Let's call word meaningful if it represents DB field descriptor or DB field value.
- AD election is represented by DB. DB **meaningful** fields (i.e. they don't represent primary or foreign keys) contain election data. Each meaningful fields has a list of descriptors. Between descriptors and meaningful fields exists an one-to-one attitude.
- The purpose of MR processing is to match MR meaningful words against the DB fields descriptors.
- To highlight the complexity of such matching, it is enough to consider quite a simple MR: "Who won an election in <constituency>?". Without knowing "who is who" and meaning of "won election" QAMEN-EMBL cannot answer this question. To explain it to QAMEN-EMBL the Production Rules (PR) need to be involved.

The subset of PR in format:

**<Precondition>  $\mapsto$  <Antecedent>  $\Rightarrow$  <Consequent>**

is shown below.

1. AD:Election  $\mapsto$  who  $\Rightarrow$  candidate
2. AD:Election  $\mapsto$  [candidate]:<win $\oplus$ won>  $\Rightarrow$  [SQL]:<MAX(votes)>
3. AD:Athletics  $\mapsto$  [runner]:<win $\oplus$ won>  $\Rightarrow$  [SQL]:<MIN(time)>
4. AD:Athletics  $\mapsto$  [shooter]:<win $\oplus$ won>  $\Rightarrow$  [SQL]:<MAX(distance)>
5. AD:Election  $\mapsto$  votes  $\Rightarrow$  [Field]:<CANDIDATE.VOTE>
6. AD:Election  $\mapsto$  candidate  $\Rightarrow$  [Field]:<CANDIDATE.[CANDIDATE NAME]>
7. AD:Election  $\mapsto$  party  $\Rightarrow$  [Field]:<CANDIDATE.PARTY>
8. AD:Election  $\mapsto$  [party]:<win $\oplus$ won>  $\Rightarrow$  [SQL]:<TOP1, SUM(votes), DESC>

where  $\oplus$  - denotes "exclusive OR".

**Precondition** consist of **class<sub>1</sub>:value<sub>1</sub> {& class<sub>i</sub>:value<sub>i</sub>}**.

**Antecedent** might be represented by: (i) **single word** (e.g. *who, won, August, seven, etc.*), (ii) **sequence of words** (e.g. *as soon as, create KB, How are you doing, etc.*), or (iii) **pair - [context]:<value>**.

Context allows one to avoid word ambiguity and thereby distinguish difference between "Candidate won an election" and "Party won an election".

Presentation of **Consequent** is similar to Antecedent structure except (iii). For *Consequent pair* represents **[descriptor]:<value>**.

For AD Election subset (1, 2, 5..8) of PR is used. PR 3 and 4 in fact show another meaning of the same word *won* but for a different AD.

- The final step of a phrased MR to SQL query conversion is quite complicated because it is necessary to access data from many different tables within an AD and join those tables together in SQL query. Result of conversion of MR "Which party won the election" to SQL-query using selected PR and executing of produced SQL query is shown in Figure 4.

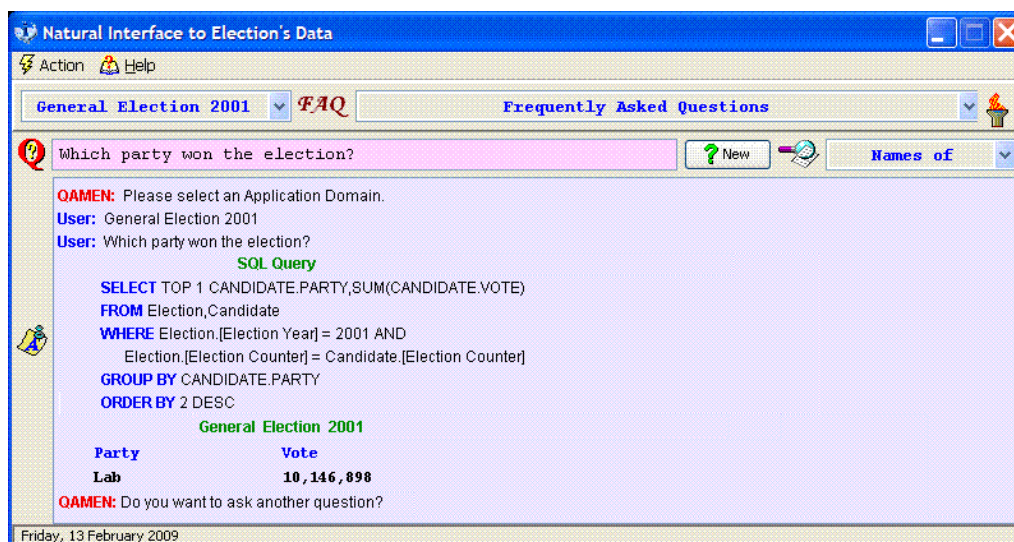


Figure 4. QAMEN-PC: Request Parsing: SQL query creation

Using QAMEN-EMBL for testing is quite expensive and that is why QAMEN-EWEB, which is oriented toward AD Election, was used. Because AD is predetermined, for QAMEN-EMBL (QAMEN-EWEB) it would be easy to understand what the MS meant by the entered MR. Instead of type in the MR "What was the result in Plymouth Devonport constituency in 2001 General Election?" it would be enough for MS to text in just "Plymouth Devonport 2001". By the way, exactly these three words represent the result of initial MR processing. After conversion to SQL query and running it result will be displayed in the mobile format (see Figure 5) i.e. QAMEN-EMBL (QAMEN-EWEB) is trying to minimize the space for response (compare with Figure 3).

## Conclusion

Like any technology, mobile telecommunication can have a wide variety of effects on political behaviour and practices, and the fact that it has been around for such a short period of time makes it impossible to reach a general conclusion about its ultimate impact. Nevertheless, we hope that in future elections, the use of mobile technology will become the standard and play an even larger role, helping to make nationwide election campaigns available to all mobile users.

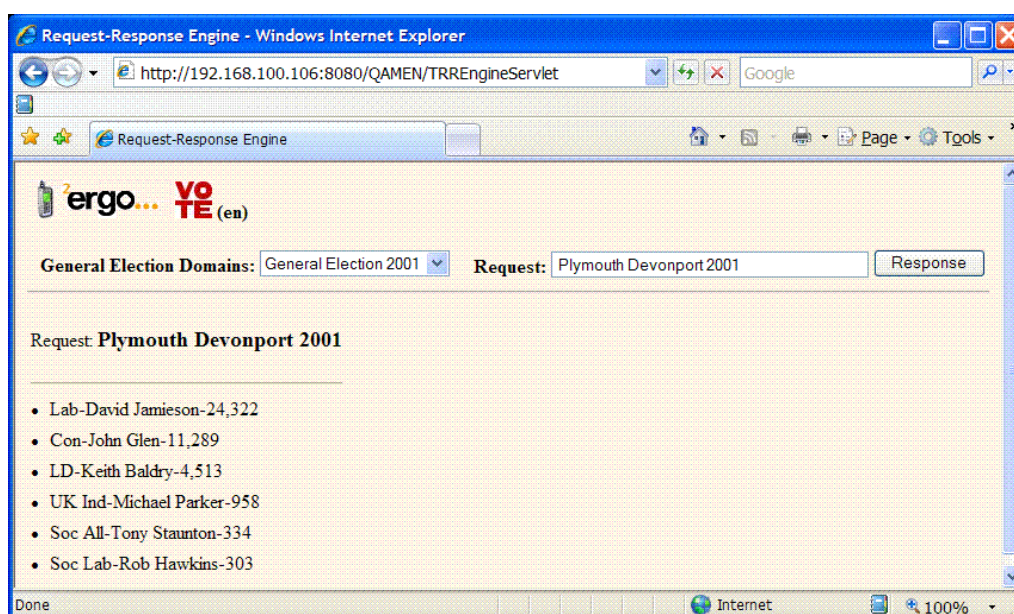


Figure 5. QAMEN-EWEB: Question-Answering interface

Without a doubt mobile phones have the potential to change certain aspects of political behaviour, including people's desire to participate in the political process. We believe that political parties should have their mobile campaign constantly, every year and not only during the three-and-a-half weeks of the election campaign every four or five years. Moreover, it would be very useful to have access to the results of any general election instantly via mobile phone. In our paper we tried to demonstrate that QAMEN-E<sub>MBL</sub> can provide such services to mobile users.

---

## Bibliography

---

- [1] Vladimir Lovitskii, Michael Thrasher, David Traynor, "Automated Response To Query System", *International Journal "Information Theories & Applications"*, Vol 15(2), 143-152, 2008.
- [2] Lee Johnston, Vladimir Lovitskii, Ian Price, Michael Thrasher, David Traynor, "Personalized Question-Answering Mobile System", International Book Series: "*Information Science and Computing*", book 2: "Advanced Research in Artificial Intelligence", 123-132, 2008.
- [3] [http://news.bbc.co.uk/1/hi/uk\\_politics/7582671.stm](http://news.bbc.co.uk/1/hi/uk_politics/7582671.stm)
- [4] [http://www.text.it/mediacentre/facts\\_figures.cfm](http://www.text.it/mediacentre/facts_figures.cfm)
- [5] Heike Hermanns, "Mobile Democracy: Mobile Phones as Democratic Tools", *Journal "Politics"*, Vol 28(2), 74-82, 2008.
- [6] Sandra L. Suarez, "Mobile Democracy: Text Messages, Voter Turnout and the 2004 Spanish General Election", *Representation*, Vol 42(2), 117-128, 2006,  
<http://www.temple.edu/polsci/suarez/documents/MobileDemocracy2006.pdf>
- [7] Sean Coughlan, "The first mobile phone election",  
[http://news.bbc.co.uk/1/hi/uk\\_politics/vote\\_2005/frontpage/4457723.stm](http://news.bbc.co.uk/1/hi/uk_politics/vote_2005/frontpage/4457723.stm)
- [8] <http://mobileactive.org/texting-it-in>
- [9] <http://www.pcpro.co.uk/news/242835/election-to-be-decided-by-mobile-phone.html>
- [10] Jonathan Spalter, "How Mobile Technologies are Changing Elections",  
[http://www.huffingtonpost.com/jonathan-spalter/how-mobile-technologies-a\\_b\\_134996.html](http://www.huffingtonpost.com/jonathan-spalter/how-mobile-technologies-a_b_134996.html)
- [11] [www.smseducation.org](http://www.smseducation.org)
- [12] Ken Braithwaite, Mark Lishman, Vladimir Lovitskii, David Traynor, "Distinctive Features of Mobile Messages Processing", *International Journal "Information Theories & Applications"*, Vol 14(2), 154-160, 2007.
- [13] T.Coles, V.A.Lovitskii, "Text Searching and Mining", *Journal of Artificial Intelligence*, National Academy of Sciences of Ukraine, Vol 3, 488-496, 2000.

---

## Authors information

---

**Elena Long** – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, e-mail: [elena.long@plymouth.ac.uk](mailto:elena.long@plymouth.ac.uk)

**Vladimir Lovitskii** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [vladimir.lovitskii@fsmail.net](mailto:vladimir.lovitskii@fsmail.net)

**Michael Thrasher** – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, e-mail: [mthrasher@plymouth.ac.uk](mailto:mthrasher@plymouth.ac.uk)

**David Traynor** – 2 Ergo Limited, 4<sup>th</sup> Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: [david.traynor@2ergo.com](mailto:david.traynor@2ergo.com)

---

## COMPUTER SUPPORT OF SEMANTIC TEXT ANALYSIS OF A TECHNICAL SPECIFICATION ON DESIGNING SOFTWARE

Alla Zaboлева-Zotova, Yulia Orlova

**Abstract:** *The given work is devoted to development of the computer-aided system of semantic text analysis of a technical specification. The purpose of this work is to increase efficiency of software engineering based on automation of semantic text analysis of a technical specification. In work it is offered and investigated a technique of the text analysis of a technical specification is submitted, the expanded fuzzy attribute grammar of a technical specification, intended for formalization of limited Russian language is constructed with the purpose of analysis of offers of text of a technical specification, style features of the technical specification as class of documents are considered, recommendations on preparation of text of a technical specification for the automated processing are formulated. The computer-aided system of semantic text analysis of a technical specification is considered. This system consist of the following subsystems: preliminary text processing, the syntactic and semantic analysis and construction of software models, storage of documents and interface.*

**Keywords:** *natural language, semantic text analysis, technical specification.*

**ACM Classification Keywords:** *I.2.7 Natural Language Processing*

**Conference:** *The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009*

---

### Introduction

---

Now designing software includes development of requirements or technical specification (TS), system or technical project (TP), programming or design engineering, trial operation, support and improvement. It is necessary to take into account interdependence of all basic parts of process of designing software from toolkit, technologies and the organizations of works. The majority of works to areas CAD systems are directed on creation and perfection of toolkit for automation of designing process. The significant contribution to development CAD systems have brought in V.I.Averchenkov, G.S.Altshuller, A.V.Andrejchikov, N.P.Buslenko, V.P.Bykov, B.S.Voinov, G.D.Volkova, V.Gasparsky, Дж. K.Dzhons, Дж. Dickson, M.F.Zaripov, V.A.Kamayev, K.V.Kumunzhiev, V.M.Kurejchik, P.M.Mazurkin, I.Muller, I.P.Norenkov, I.J.Petrova, A.I.Polovinkin, A.F.Pohilko, J.M.Solomentsev, F.Hanzen, P.Hill, A.Holl, etc.

Most known of the commercial software products used at designing of the software, basically are intended for visualization intermediate and end results of process of designing. Some of them allow to fully automate last design stages: generation of a code, creation of the accounting and accompanying documentation, etc. Thus the problem of automation of the initial stage of designing - formations and the analysis of the text of the technical project remains open. It is connected to extraordinary complexity of a problem of synthesis and the analysis of semantics of the technical text for which decision it is necessary to use methods of an artificial intellect, applied linguistics, psychology, etc. However, it is possible to come nearer to achievement of the given purpose, having allocated some small subtasks quite accessible to the decision by known methods of translation.

Proceeding from the aforesaid, it is possible to draw a conclusion, that the problem of creation of means for automation of process of designing is actual [1].

Ideas of a developed direction realization of the unified procedures of the designing equally answering to requirements of the expert - designer and requirements to technology to modelling of software products is main.



The purpose of this work is to increase efficiency of software engineering based on automation of semantic text analysis of a technical specification.

To achieve this purpose it is necessary to solve the following tasks:

1. To carry out the analysis of software engineering process and models of semantic text analysis;
2. To develop a technique of the text analysis of a technical specification;
3. To develop and investigate semantic model of the text of a technical specification;
4. To develop algorithmic maintenance of analysis of text of a technical specification and automatic construction of the software models;
5. To realize developed formalisms, a technique and algorithms as system of automation of the initial stage of designing software.

### A Technique Of The Text Analysis Of A Technical Specification

In work it is offered and investigated a technique of the analysis of the text of a technical specification is submitted, the fuzzy attribute grammar of a technical specification, intended for formalization of limited Russian is constructed with the purpose of analysis of offers of text of a technical specification, style features of the technical specification as class of documents are considered, recommendations on preparation of text of a technical specification for the automated processing are formulated.

A technique of the analysis of the text of a technical specification consist of three stages: semantic text processing, creation of frame structure and creation of data flow diagrams of system described in the technical specification. (see Figure 1).

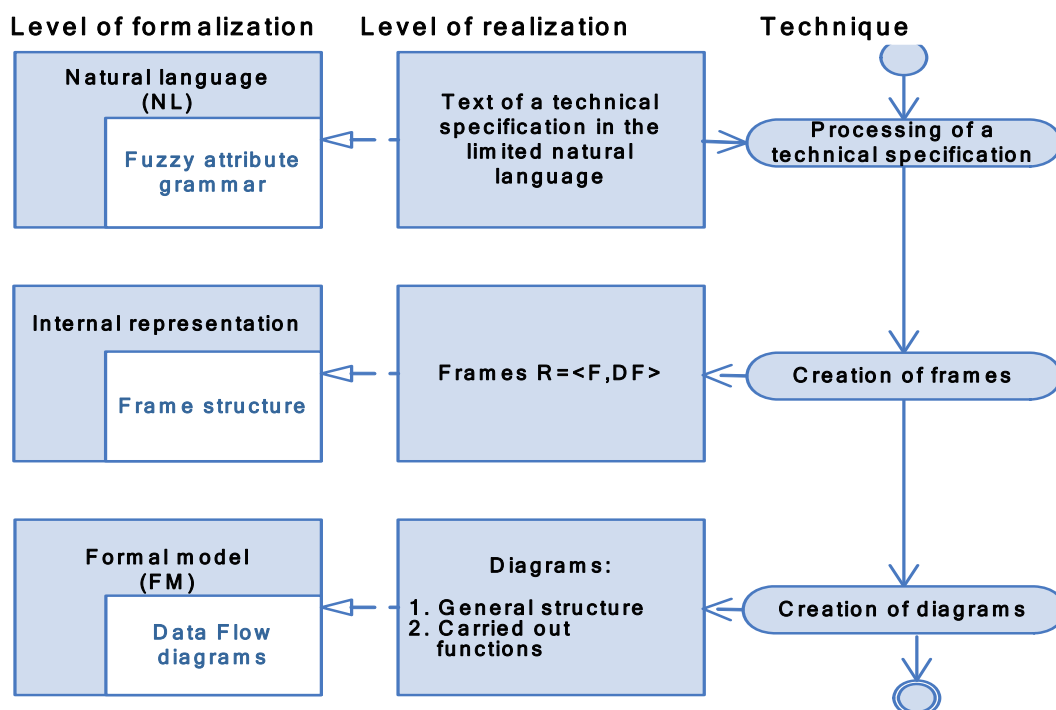


Figure 1: Technique Of The Text Analysis Of A Technical Specification

For realization of the first stage of a technique the semantic model of the text of a technical specification, including the requirements formulated as the document in the limited natural language has been developed; the

second stage - the frame structure being internal representation of requirements; the third stage - model of software as the description of requirements in graphic language Data Flow Diagrams.

The semantic model of the text of a technical specification contains the developed expanded fuzzy attribute grammar above frame structure of the formal document "Technical specification" which allows to display contents TS most full.

The expanded fuzzy attribute grammar, necessary for the automated analysis of the text of a technical specification, is determined as:

$$AG = \langle N, T, P, S, B, F, A, D(A) \rangle,$$

where N - final set of non-terminal symbols; T - not crossed with N set of terminal symbols; P - final set of rules; S - the allocated symbol from N, named an initial symbol; B - set of linguistic variables  $\beta_{k,i}$ , corresponding to terminal symbols T (a variable i on k level); F - set of functions of a belonging  $f_{k,i}$ , determining a degree of belonging  $m_{k,i}$  linguistic variables  $\beta_{k,i}$ ; A - set of attributes,  $A = A_{\text{sin}} \cup A_{\text{sem}}$ , where  $A_{\text{sin}}$  - syntactic attributes,  $A_{\text{sem}}$  - semantic attributes; D (A) - final set of semantic actions. The fragment of grammar is submitted in table 1.

Table 1: Fragment of the developed fuzzy attribute grammar above frame structure of a technical specification

$\beta_1$	<i>&lt;list of incoming data flows &gt;</i>	<i>&lt;incoming data flow name &gt; :: 'Name' &lt;incoming data flow description&gt; :: 'Contents' &lt; list of incoming data flows &gt;   <math>\epsilon</math></i>
	<i>&lt;incoming data flow description&gt;</i>	<i>The text containing "entrance" or "entrance data" :: 'Clause' &lt;incoming data flow&gt;::"Frame Data Flow=Creation", "Input=Giving"</i>
$\beta_{1,2}$	<i>&lt;incoming data flow&gt;</i>	<i>[&lt;Number of data units&gt;]:: "Slot AMOUNT OF DATA = Giving" [&lt;Type of data&gt;]:: "Slot TYPE OF DATA = Giving" &lt;the Name of incoming data flow &gt;:: "Slot NAME OF INCOMING DATA FLOW = Giving"</i>
$\beta_2$	<i>&lt;function specification &gt;</i>	<i>&lt;function type &lt;name of the functions liss&gt;:: 'Name'&lt;function description&gt;:: "Frame FUNCTION = Creation"; &lt; List of functions &gt;   <math>\epsilon</math></i>
$\beta_{2,1}$	<i>&lt; function type &gt;</i>	<i>«main»   «basic»   «additional»</i>
$\beta_{2,2}$	<i>&lt;function description &gt;</i>	<i>&lt;Name of function&gt;:: 'Name', "Slot NAME OF FUNCTION = Giving" &lt;List of incoming data flow&gt; &lt;List of outgoing data flow&gt;</i>

Linguistic variables from set  $B = \{\beta_{k,i}\}_{k,i}$  used for the analysis of the text of a technical specification is described by the following five:

$$\beta_{k,i} = \langle \beta, T(\beta), U, G, M \rangle,$$

$\beta$  - name of linguistic variable (basis for development, purpose of development, technical requirements to a program product, a stage and development cycles, etc.);

$T(\beta)$  - language expressions. For linguistic variables of the top level they are the linguistic variables corresponding to terminals of the right part of a rule. For linguistic variables of the bottom level – fuzzy variables, that is expressions of a natural language.

U - Set of all probable values,  $T(\beta) \subset U$ ;

G - rules of the morphological and syntactic description of language expressions which determine syntactic attributes  $A_{\text{sin}}$ ;

M - a semantic rule for linguistic variables which is induced by morphological and syntactic rules as the sense of a term in T is in part determined by its syntactic tree, and semantic attributes Asem.

Methods of representation connections between rules are broadcast on language of fuzzy mathematics. Thus connections are represented by fuzzy relations, predicates and rules, and sequence of transformations of these relations - as process of an fuzzy conclusion.

Linguistic variables of the top level are compound, that is include linguistic variables of the bottom level. Due to this it is possible to construct a tree of linguistic variables and to establish dependence between them.

Functions of an a belonging from set  $F = \{fk,i\}$ ,  $i$  linguistic variables  $\{\beta k,i\}$ ,  $k,i$ , are necessary for construction of an fuzzy conclusion. In particular, to each rule of grammar from set P function of a belonging  $fk,i$  is put in conformity. This dual system of substitutions is used for calculation of sense of a linguistic variable.

Actually grammar of a technical specification is used for splitting the initial text of the document into sections and processings of most important of them for our problem. It needs precise observance of structure of the document. Technical specification represents the structured text consisting of sequence of preset sections.

The frame structure of the technical specification is submitted as:

$$R = \langle N_R, \overline{F}_R, \overline{I}_R, \overline{O}_R \rangle$$

where  $N_R$  is a name of system,  $F_R$  is system functions vector,  $I_R$  is incoming data flows vector,  $O_R$  is outgoing data flows vector.

$$\overline{F}_R = \langle F_R^1, F_R^2, \dots, F_R^k \rangle, \text{ then } F_R^i = \langle N_F^i, \overline{I}_F^i, D_F^i, G_F^i, H_F^i, \overline{O}_F^i \rangle,$$

Where  $N_F^i$  - a name of function  $F_R^i$ ,  $\overline{I}_F^i$  - incoming data flows vector of F function,  $D_F^i$  - the name of the action which are carried out by function,  $G_F^i$  - subject of the function action,  $H_F^i$  - restrictions on function,  $\overline{O}_F^i$  - a outgoing data flows vector of F function.

Let's denote the data flow by DF (Data Flow), then  $I_R, O_R, I_F, O_F$  are denoted by:

$$DF = \langle N_{DF}, D_{DF}, T_{DF}, C_{DF} \rangle$$

Where  $N_{DF}$  - data flow name,  $D_{DF}$  - data flow direction,  $T_{DF}$  - data type in flow,  $C_{DF}$  - data units per frame.

The model proposed is represented as a frame model with "a-kind-of" links (see Figure 2).

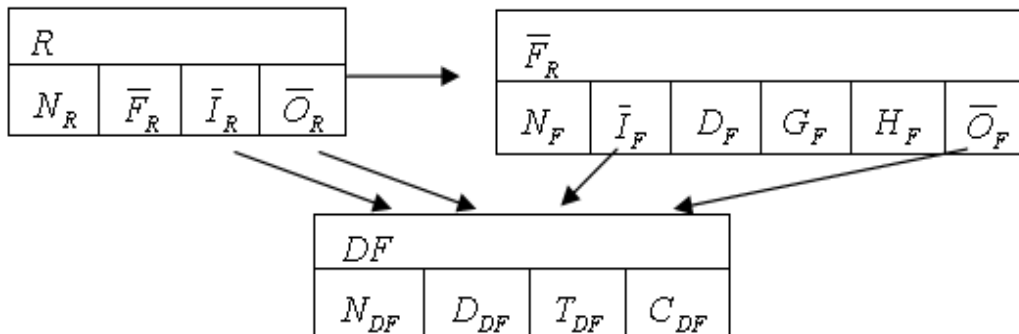


Figure 2: Frame network



### Computer-Aided System Of Semantic Text Analysis Of A Technical Specification

The computer-aided system of semantic text analysis of a technical specification consists of the following subsystems: preliminary text processing, the syntactic and semantic analysis and construction of software models, storage of documents and interface (see Figure 3).



Figure 3: Architecture of computer-aided system of semantic text analysis of a technical specification

Preliminary text processing is necessary to share of a technical specification on separate lexemes. The incoming information of a subsystem is the text of a technical specification in the limited natural language, the target information - tables of sections, sentences and lexemes of a considered technical specification. Results can be submitted both as corresponding tables, and as a tree of sections.

Already after the first stage work not with the text of a technical specification, but with its parts submitted on sections is made. On a course of work of a technical specification shares all over again on more and more fine sections, then on separate sentences (with preservation of sections structure) and lexemes with the instruction of an accessory to sentences.

Preliminary text processing is carried out with use of final automatic device. During the work of final automatic device the symbols acting on its entrance, collect in the buffer. In the certain conditions of final automatic device record of the current contents of buffer in one of tables then the buffer is devastated is carried out. Work of automatic device proceeds up to achievement of a final condition.

After that the received tables act on an entrance of a subsystem of syntactic and semantic analysis. The semantic analysis of a text is made on the basis of the developed grammar of a text of technical specification.

Rules of top level serve for analysis of sections of top level. Rules for analysis of sections consist of two parts: the first part serves for analysis of a section name; the second part serves for analysis of a text contents in section. Symbols of the given grammar possess syntactic attributes. In attributes of non-terminal symbols names of frames or names of slots in which the information received during the further analysis should be placed are specified. Syntactic attributes of text can be in addition specified in attributes of terminal symbols. Comparison of words at analysis is made in view of their morphology. During analysis the syntactic and morphological analysis are made only in the event that there is such necessity that time of performance of semantic analysis is considerably reduced.

Let's consider a fragment of the developed attribute grammar submitted in a xml-format:

```
... <global-rule id="Section42" comment = "Section 4.2. Requirements to functional characteristics">
<rule><ruleref uri="#Section42Name"/><ruleref uri="#Section42x"/></rule></global-rule>
<global-rule id="Section42Name" sectionPart="Name" comment= "Heading of the unit 4.2."><rule><clause
clauseType="UNCERTAIN"/><rule type="or"><words contains="Functions"/> <words contains= " functional
characteristics "/> </rule></rule></global-rule>
<global-rule id="Section42x" frame= "FunctionFrame" frameSlot="Function" comment="Function"><rule> <ruleref
uri="#Section42xName" /><ruleref uri="#Section42xContent" /> </rule></global-rule>
<global-rule id="Section42xContent" sectionPart="Content" comment="Inputs and outputs of
function"><rule><ruleref uri= "#Section42xInputs" minOccurs="0"/><ruleref uri="#Section42xOutputs"
minOccurs="0"/></rule></global-rule>
<global-rule id="Section42xInputs" comment="Inputs of function">
<rule><sentence/><clause/><rule type="or"><words contains="Inputs"/> <words contains="entrance
data"/></rule><ruleref uri="#Input" maxOccurs="unbounded"/></rule></global-rule> ...
```

The morphological and syntactic modules used in the program, are modules of the foreign developer. If in a rule of grammar there is a terminal having syntactic attribute the mechanism of syntactic analysis for current sentences is started [2].

After creation of a tree of analysis construction of frame description of a technical specification begins. For this purpose the information on frames and names of slots which contains in attributes of symbols of grammar is used.

The received frame structure contains the significant information about system: data about inputs and outputs of system, functions and restrictions. For each function inputs and outputs also are allocated. It allows to receive data flow diagrams of system which is described in a technical specification on the basis of frame structure.

The subsystem "Construction of data flow diagrams" carries out construction and ordering the column of data flows, and also creation the figures of data flow diagrams in Microsoft Office Visio.

For construction of data flows it is prospected of functions inputs conterminous to system inputs. Then functions on which all inputs data act, are located on the one level of diagram. Their inputs incorporate to system inputs. Further it is prospected functions which inputs coincide with outputs of functions received on the previous step. They are located on the following level, their inputs incorporate to outputs of the previous levels functions and with system inputs.

Work of algorithm proceeds until all functions will not be placed on the diagram. After that connection of function outputs with necessary system outputs is made.

---

The computer-aided system of semantic text analysis of a technical specification is developed on Microsoft .NET Framework 2.0 platform (language of development C#) using integrated development environment Visual Studio 2005.

---

### Scientific Novelty

---

Scientific novelty consists in the following: a technique of text analysis of a technical specification at the initial stages of software engineering, including semantic model of text of a technical specification, transformation matter of text into the frame structure and construction of model of the software on its basis are developed.

---

### Practical Value

---

Practical value of work is that as a result of development and introduction of a suggested technique quality of software engineering raises due to automation of routine work of the person on extraction of helpful information from standard documents and to displaying it as software models.

---

### Conclusions and Future Work

---

Software designing differs from designing in other areas of a science and technics a little, therefore it is possible to expand results of the given work for application in other areas of human knowledge. Thus, opening prospects raise a urgency of the given work.

---

### Bibliography

---

1. Kamsay, A. Computer-aided syntactic description of language systems/ A. Kamsay// Computational linguistics. An international handbook on computer-oriented language research and applications.- Boston: Walter de Gruyter, 1989.- P.204-218
2. Reyle, U. Natural language parsing and linguistic theories/ U. Reyle.- Berlin: Rohrer Dordrecht, 1998.- 625 p.
3. Tools Development For Computer Aided Software Engineering Based On Technical Specification's Text Analysis / A.Zaboleeva-Zotova, Y.Orlova // Interactive Systems And Technologies: The Problems Of Human-Computer Interaction: Proc. of the Int. Conf., Ulyanovsk.

---

### Authors' Information

---

**Alla V. Zaboleeva-Zotova** – PhD, professor; CAD department, Volgograd State Technical University, Lenin av., 28, Volgograd, Russia; e-mail: [zabzot@vstu.ru](mailto:zabzot@vstu.ru)

**Yulia A. Orlova** – PhD; CAD department, Volgograd State Technical University, Lenin av., 28, Volgograd, Russia; e-mail: [yulia.orlova@gmail.com](mailto:yulia.orlova@gmail.com)

---

## К АНАЛИЗУ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ОБЪЕКТОВ

**Александр Палагин, Сергей Кривый, Виталий Величко, Николай Петренко**

**Аннотация:** Рассматриваются проблемы анализа естественно-языковых объектов (ЕЯО) с точки зрения их представления и обработки в памяти компьютера. Предложена формализация задачи анализа ЕЯО и приведен пример формализованного представления ЕЯО предметной области.

**Ключевые слова:** термины предметной области, формальная логическая система, онтология.

**ACM Classification Keywords:** I.2.4 Knowledge Representation Formalisms and Methods - Representation languages, I.2.7 Natural Language Processing - Language models

**Conference:** The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009

---

### Введение

---

Проблема извлечения знаний из естественно-языковых объектов, их представление и обработка относятся к области искусственного интеллекта. Этой чрезвычайно важной проблемой занимались многие выдающиеся ученые (Аристотель, Лейбниц, Тарский и др.). В общей проблеме анализа и понимания ЕЯО, во всем многообразии работ [1-5], методов, методологий и терминологии явно просматриваются следующие задачи:

- синтактико-семантический анализ ЕЯО,
- анализ ЕЯО с целью выбора подходящего представления в памяти компьютера и дальнейшей эффективной обработки,
- разработка методов извлечения знаний из ЕЯО.

Рассмотрим эти задачи более детально.

Проблема анализа ЕЯО естественным образом связана с **операционной и денотационной семантикой** языка. В системах искусственного интеллекта и, в частности, в системах анализа ЕЯО преобладает операционная семантика, так как очевидно преобладание интереса к средствам представления знаний (**как?**), а не к их содержанию (**что?**), которое подлежит формализации.

Системы (языки) представления знаний определяют конструкции, поддерживаемые компьютером и обеспечивающие простоту организации процедур, и как следствие – максимальную унификацию языковых средств.

На информационно-логическом уровне преследуется совершенно иная цель - обеспечение отражения содержания объектов реального мира в языковых конструкциях, согласующихся с восприятием, свойственным человеку. Для пользователя такой системы (схемы) компьютерное представление неудобно, поскольку он общается (или, по крайней мере, должен) с системой на естественном языке.

Однако, для системного аналитика, использующего язык формализации для построения модели предметной области (ПдО) и создающего для этого ориентированные на эту ПдО словари понятийных единиц и на их основе - интеллектуальные системы, наполненные конкретным знанием, компьютерное представление играет ключевую роль. Аналитика приходится иметь дело со словами (терминами) и конструкциями какой-то части естественного языка, именно им давать толкование в терминах языка формализации, устанавливать между ними информационно значимые связи. Поэтому, в случае

расхождения языка формализации со строем мышления специалиста, реализация системы обработки данных в конкретном смысловом материале может стать слишком сложной или вообще невозможной для решения проблемой.

Существует много языков представления знаний, которые формализуют ЕЯО фрагментарно, а иногда и недостаточно четко. Поэтому наиболее приемлемым на сегодняшний день являются языки логического типа, в которых понятие "связь по смыслу" формализуется с максимальной полнотой. Это подтверждается и практикой. Одним из главных вопросов при этом является вопрос об уровне абстрактности зафиксированных категорий логического языка. Если нас интересует лишь общая схема логического следования, то достаточно рассматривать лишь исчисление предикатов первого порядка с правилом резолюций в качестве правила вывода, не конкретизируя ничего, кроме списка аксиом и логических констант. Однако, при более богатом содержанием ПдО должен быть решен вопрос о типах термовых констант и виде нелогических аксиом (собственных аксиом), используемых при формализации знаний об эмпирических объектах.

Из всего сказанного следует проблема **разработки языка логического типа для представления знаний, содержащихся в тексте**. Очевидно, что этот язык должен быть расширяемым в сторону способности более полного описания объектов, способности делать предсказания (гипотезы), учитывающие контекст.

---

### Формальная постановка задачи анализа ЕЯО

---

Пусть  $T = t_1, t_2 \dots t_n$  естественно-языковой текст в алфавите  $X$ , т. е.  $T \in L(Y)$ , где  $L(Y)$  - язык над алфавитом  $Y$ , а  $t_i \in T$  - предложения,  $i = 1, 2, \dots, n$ .

Каждое предложение  $t_i \in T$ , в свою очередь, имеет структуру  $t_i = t_{i_1}, t_{i_2} \dots t_{i_m}$ , где  $t_{i_j}$  содержательно означают грамматические единицы, из которых построено предложение  $t_i$ . Если  $t_{i_j} \in t_i$ , то  $C_L(t_{i_j}) = t_{i_1} \dots t_{i_{j-1}}$  и  $C_R(t_{i_j}) = t_{i_{j+1}} \dots t_{i_m}$  будем называть левым и правым контекстом слова  $t_{i_j}$  соответственно.

С текстом  $T$  свяжем такие объекты:

- $S$  – словарь языка  $L(Y)$ , где содержатся слова  $t_{i_j}$  со своими определителями-  $a_i$  и  $\tau_i$ ;
- $\gamma \subseteq T \times S$  – отношение, определяющее возможные значения и типы слова в словаре  $S$ ;
- $M = (D, \Pi)$  – предметная модель, на которой интерпретируется текст  $T$  из предметной области  $D$ ;
- $\phi \subseteq T \times M$  – отношение интерпретации текста  $T$  на модели  $M = (D, \Pi)$ .

Сигнатура предикатов  $\Pi = \{\pi_1^{k_1}, \dots, \pi_r^{k_r}\}$  содержит атомарные предикаты, из которых можно строить сложные формулы. Сейчас мы не будем фиксировать эту сигнатуру, поскольку она зависит от предметной модели  $M$ . Поскольку модель не уточняется, то и сигнатуру уточнить нельзя. Заметим только, что каждый атомарный предикат имеет тип.

Определим теперь правила вычисления отношений  $\gamma$  и  $\phi$ .

Отношение  $\gamma$  имеет достаточно простой способ вычисления:

$$\gamma(t_{i_j}) = \{(a_1, \tau_1), (a_2, \tau_2), \dots, (a_s, \tau_s)\},$$

где  $a_i$  – возможные значения слова  $t_{ij}$ ,  $\tau_i$  – его возможные типы (например, часть речи). Может случиться, что  $\gamma(t_{ij}) = \emptyset$ . В этом случае значение этого слова считается неопределенным и требует пополнения словаря  $S$ .

Отношение  $\phi$  определяется следующим образом:

$$\phi(T) = \phi(t_1) \dots \phi(t_n),$$

где  $\phi(t_i) = \left\{ \phi(\gamma(t_i)\gamma(C_R(t_i))), \phi(\gamma(C_L(t_i))\gamma(t_i)\gamma(C_R(t_i))), \dots, \phi(\gamma(C_L(t_i))\gamma(t_i)) \right\}$ ,

$\phi(\gamma(t_{ij})) = \gamma(\phi(t_{ij})); \phi(\gamma(C_L(t_{ij}))) = C_L(\phi(\gamma(t_{ij}))); \phi(\gamma(\pi_r^k(p_1, \dots, p_k))) = \gamma(\phi(\pi_r^k))(\phi(\gamma(p_1), \dots, \gamma(p_k)))$ ,

где  $\gamma(\phi(\pi_r^k))$  имя предиката, тип которого согласован с аргументами  $\gamma(p_1), \dots, \gamma(p_k)$ . Из этой формальной постановки проблемы анализа ЕЯО следует, что основные задачи анализа сводятся к таким:

- конкретизировать предметную модель  $M$ ; эта задача основная в связи с тем, что предметная модель является по существу базой знаний (конкретизация состоит в том, чтобы определиться с формальным логическим языком, правилами вывода, аксиоматикой и т.п.);
- показать вычислимость отношений  $\gamma$  и  $\phi$  на предметной модели  $M$ ;
- построить алгоритмы вычисления отношений  $\gamma$  и  $\phi$ ;
- при вычислении отношений  $\gamma$  и  $\phi$  контролировать соответствие типов аргументов и предикатов;
- определить взаимодействие алгоритмов вычисления  $\gamma$  и  $\phi$  с системами синтаксического и семантического анализа текста.

Рассмотрим пример использования введенного формализма для анализа ЕЯ текста. Задача состоит в том, чтобы получить формальное представление текста в полуавтоматическом режиме взаимодействия с пользователем – экспертом предметной области, при минимальном его участии в процессе анализа.

---

### Использование предложенного формализма для анализа ЕЯО

---

Словарь  $S$ , со словами  $t_{ij}$  и их определителями  $a_i$ ,  $\tau_i$  создается на основе результатов лексикографических исследований [4] и, по-возможности, является универсальным, независимым от предметной области. В качестве примера ЕЯ текста выберем небольшой фрагмент из предметной области "сверхтвердые материалы" [6]. Слева от предложений указан номер строки в фрагменте текста.

- 1 ПОЛУЧЕНИЕ И СВОЙСТВА ПОЛИКРИСТАЛЛИЧЕСКИХ МАТЕРИАЛОВ НА ОСНОВЕ АЛМАЗА
- 2 1.Материалы, синтезированные из неалмазного углерода
- 3 Поликристаллические материалы на основе алмаза по своим физико-механическим свойствам
- 4 могут быть близкими к монокристаллам, а по некоторым — и превосходить их. Так,
- 5 большинство алмазных поликристаллов обладает изотропией свойств, отличается высокой
- 6 износостойкостью и превосходит монокристаллы по трещиностойкости (ударной вязкости).
- 7 По принятой в настоящее время классификации поликристаллические сверхтвердые
- 8 материалы (ПСТМ) делятся на четыре группы.
- 9 I группа — ПСТМ, получаемые при высоких давлениях и температурах путем превращения
- 10 графита в алмаз в присутствии специальных сплавов-растворителей углерода. К этой
- 11 группе относятся, например, поликристаллы типа баллас (АСБ) и карбонадо (АСПК).
- 12 Следует отметить, что в месторождениях алмазов встречаются природные
- 13 поликристаллы подобных типов.
- 14 II группа — ПСТМ, получаемые путем спекания алмазных порошков в условиях высоких

- 15 статических давлений и температур.  
 16 III группа — ПСТМ, получаемые при высоких статических давлениях и температурах на  
 17 подложке из твердого сплава или стали, а также ПСТМ в твердосплавной или металлической  
 18 обойме.  
 19 IV группа — ПСТМ, получаемые методом спекания алмазных порошков при более низких  
 20 давлениях (до 1, 9 ГПа) в присутствии связующего материала.

Уточним сигнатуры предикатов предметной модели. С помощью поверхностного синтактико-семантического анализа [7] для всех слов из приведенного примера найдем  $\gamma(t_i)$ , используя словарь  $S$ .

Для каждого предложения текста строятся деревья зависимостей, в которых определяются синтактико-семантические отношения между словами. Далее автоматически выделяются из текста термины предметной области  $D$  в соответствии с заданными шаблонами, например: аббревиатура, существительное, прилагательное\_существительное, существительное\_прилагательное\_существительное и типами синтактико-семантических отношений: объектное, принадлежность (между двумя существительными), определительное (между прилагательным и существительным), однородные слова. В качестве терминов, используемых в формальном представлении текста, автоматически выбираются преимущественно многословные словосочетания, в которых значения существительных уточняются с помощью связанных прилагательных.

Приведем фрагмент списка терминов, состоящего из отдельных слов и словосочетаний: *свойства поликристаллических материалов, алмаз, материал, неалмазный углерод, поликристаллический материал, физико-механическое свойство, алмазный поликристалл, поликристалл, изотропия свойств, высокая износостойкость, трещиностойкость, ударная вязкость, поликристаллический сверхтвердый материал, ПСТМ, I группа ПСТМ, II группа ПСТМ, III группа ПСТМ, IV группа ПСТМ, высокое давление, давление, высокая температура, температура, превращение графита, графит, сплав-растворитель углерода, баллас, АСБ, карбонадо, АСПК, природный поликристалл, алмазный порошок, спекание алмазного порошка, высокое статическое давление, подложка из твердого сплава, подложка из стали, металлическая обойма, твердосплавная обойма, низкое давление, связующий материал.*

Используя словарь  $S$ , в предметной области  $D$  определяются отношения  $\phi$  для  $t_i$  с грамматическими характеристиками: глагол, существительное, причастие, деепричастие. Для этого используется информация из лексикографической базы данных. Для более точного выбора значения  $\phi$  учитываются  $C_R(t_i)$  и  $C_L(t_i)$ . Под правым и левым контекстом слова понимаются слова непосредственно связанные с  $t_i$  синтактико-семантическими отношениями. Так для словосочетаний *получаемые путем* (строки 9,14 примера), *получаемые методом* (строка 19) отношение  $\phi$  принимает значение *способ получения*, для *получаемые при*, *получаемые в присутствии* (строки 9,10) –  $\phi = \{\text{условие получения}\}$ , для *синтезированные из* (строка 2) –  $\phi = \{\text{состав}\}$ . В словосочетании *получаемые на подложке* (строки 16,17) предлог *на* в правом контексте слова *получаемые* указывает на предложный падеж, т.е. отношения места, поэтому для *получаемые на*  $\phi = \{\text{способ получения}\}$ . Для слова *делятся* (строка 8) характеристикой ( $a_i$ ) является использование его в описании отношений "класс-подкласс" (гипонимия) или "часть-целое" (меронимия). Конкретизируем значение *делятся*, используя выделенные синтактико-семантические отношения,  $C_R(t_i)$  и  $C_L(t_i)$ . Найдя в правом контексте слова *делятся* связанные слова – *на группы* (для группы  $a_i = \{\text{"класс-подкласс"}\}$ ), определяем конкретное значение  $\phi = \{\text{"класс-подкласс"}\}$ .

На основе сформированного списка терминов, уточненных значений  $\phi$  и построенных деревьев зависимостей для предложений текста, определяются элементы множества  $D$ : ГРУППА

ПОЛИКРИСТАЛЛИЧЕСКИХ СВЕРХТВЕРДЫХ МАТЕРИАЛОВ, СПОСОБ ПОЛУЧЕНИЯ, УСЛОВИЕ ПОЛУЧЕНИЯ, СОСТАВ, СВОЙСТВА. В приведенном примере в множество  $D$  попали элементы, характеризующие иерархические и атрибутивные отношения. Это обусловлено видом ЕЯО – научная проза и значениями интерпретации отношения  $\phi$ , определенными для слов из словаря  $S$ . Сформированный список элементов  $D$  уточняется экспертом предметной области.

Из построенных деревьев зависимостей фраз текста получим множество предметных констант и их семантические значения: ГРУППА ПОЛИКРИСТАЛЛИЧЕСКИХ СВЕРХТВЕРДЫХ МАТЕРИАЛОВ = {I группа, II группа, III группа, IV группа}; СПОСОБ ПОЛУЧЕНИЯ = {спекание алмазного порошка, превращения графита, металлическая обойма, твердосплавная обойма, подложка из твердого сплава, подложка из стали}; СВОЙСТВА = {высокая износостойкость, изотропия свойств}; УСЛОВИЕ ПОЛУЧЕНИЯ = {низкое давление, высокое давление, сплав-растворитель углерода, высокое статическое давление, высокая температура, связующий материал}; СОСТАВ = {неалмазный углерод}.

На следующем шаге заменим указательное местоимение *этой* (строка 10 примера), связанное со словом *группа*, на его значение в соответствии с алгоритмом, приведенным в [7]. Из предыдущего предложения (строка 9) можно получить однозначное значение данного местоимения – *I группа ПСТМ*.

В приведенном примере имеется синонимия терминов: ПСТМ, поликристаллические материалы на основе алмаза, поликристаллические сверхтвердые материалы, алмазные поликристаллы; трещиностойкость – ударная вязкость; баллас – АСБ; карбонадо – АСПК. Для задания отношения синонимии требуется просмотр списка возможных синонимов экспертом предметной области. Список возможных синонимов формируется автоматически, используя простые правила, например, аббревиатура в скобках может быть синонимом однословного или многословного термина, стоящего в предложении впереди аббревиатуры. После уточнения отношения синонимии любой из терминов синонимов может быть использован в дальнейших операциях с текстом. В строке 13 встречается неопределенное словосочетание *подобных типов*. Выделим данное словосочетание и предъявим его эксперту для уточнения значения слова *подобных*.

На подмножествах  $D$  определим отношения-предикаты: отношение «ГРУППА ПСТМ – СПОСОБ ПОЛУЧЕНИЯ» –  $R_{г-сп} \subseteq \text{ГРУППА ПСТМ} \times \text{СПОСОБ ПОЛУЧЕНИЯ}$ .

Семантическое значение  $R_{г-сп}$  в данном примере следующее:

$$R_{г-сп} = \{(I \text{ группа ПСТМ, превращение графита}), \\ (II \text{ группа ПСТМ, спекание алмазного порошка}), \\ (III \text{ группа ПСТМ, подложка из твердого сплава, подложка из стали, металлическая обойма, твердосплавная обойма}), \\ (IV \text{ группа ПСТМ, спекание алмазного порошка})\}$$

Поверхностный формальный анализ полученных отношений показывает их семантическую неоднородность. Так для III группы ПСТМ в терминах, обозначающих способ получения, не содержатся отглагольные существительные, в отличие от других предметных констант СПОСОБА ПОЛУЧЕНИЯ. Это говорит о том, что или модель  $M$  неполна и необходимо пополнять сигнатуру предикатов  $\Pi$ , или пополнять отсутствующими понятиями множество  $D$ . Кроме того, такой формальный анализ позволяет обратить внимание эксперта на точность формулировок текста предметной области.

Отношение «ГРУППА ПСТМ – УСЛОВИЕ ПОЛУЧЕНИЯ» –  $R_{г-уп} \subseteq \text{ГРУППА ПСТМ} \times \text{УСЛОВИЕ ПОЛУЧЕНИЯ}$

$$R_{г-уп} = \{(I \text{ группа ПСТМ, высокое давление, высокая температура, сплав-растворитель углерода}), \\ (II \text{ группа ПСТМ, высокое статическое давление, высокая температура}), \\ (III \text{ группа ПСТМ, высокое статическое давление, высокая температура}), \\ (IV \text{ группа ПСТМ, низкое давление, связующий материал})\}$$

Эти и другие отношения можно получить, исходя из области  $D$ , используя операции реляционной алгебры. Аналогичным способом можно получить и другие интересующие пользователя отношения-предикаты, т.е. полностью построить сигнатуру исходного множества предикатов  $\Pi$  на области  $D$ .



Покажем использование, полученных в результате автоматического анализа текста, множества предметных констант и отношений-предикатов для построения онтологии предметной области.

Практически в любой ПдО знания делятся на "статическую" и "динамическую" составляющие, а соответствующая онтология ПдО представляется онтологией объектов и онтологией процессов. Под онтологией объектов будем понимать кортеж множеств  $O = \langle X, R, F \rangle$ , где  $X$  – конечное множество концептов (понятий) заданной предметной области,  $R$  – конечное множество отношений между концептами  $X$ ,  $F$  – конечное множество функций интерпретации, заданных на множествах  $X$  и/или  $R$ . Причем для приведенного примера достаточно рассмотреть частный случай, когда множество  $F$  тождественно множеству аксиом  $A$ , представляющих истинные высказывания о соответствующих понятиях  $X$ . Аналогичную структуру примем и для онтологии процессов ПдО.

Разделим список терминов, учитывая элементы множества  $D$ , на понятия-объекты, свойства понятий-объектов, понятия-процессы и свойства понятий-процессов. Результат такого деления приведен в Таблице 1.

Таблица 1.

Понятия-объекты	Свойства понятий-объектов	Понятия-процессы	Свойства понятий-процессов
1.1. Поликристаллический сверхтвёрдый материал	1. Свойства поликристаллических материалов	1.1. Способ получения ПСТМ 1.2. Процесс получения ПСТМ	1. Условие получения
1.2. Поликристаллический материал			2. Высокая температура
1.3. Алмазный поликристалл	2. Физико-механическое свойство	2. Процесс получения ПСТМ I группы	3. Высокое давление
1.4. ПСТМ			4. Высокое статическое давление
2. Температура	3. Высокая износостойкость	3. Процесс получения ПСТМ II группы	
3. Давление			
4. Алмаз	4. Изотропия свойств	4. Процесс получения ПСТМ III группы	
5. Алмазный порошок			
6. Природный поликристалл	5.1. Трещиностойкость	5. Процесс получения ПСТМ IV группы	
7. Неалмазный углерод			
8. Графит	5.2. Ударная вязкость	6. Спекание алмазного порошка	
9. I группа ПСТМ			
10. II группа ПСТМ			
11. III группа ПСТМ			
12. IV группа ПСТМ			
13.1. Баллас			
13.2. АСБ			
14.1. Карбонадо			
14.2. АСПК			
15. Подложка из твёрдого сплава			
16. Подложка из стали			
17. Твёрдосплавная обойма			
18. Металлическая обойма			
19. Сплав-растворитель углерода			
20. Связующий материал			

Множество  $A$  сформируем из свойств понятий-объектов и свойств понятий-процессов. Полученные аксиомы приведены ниже.

Аксиомы для понятий-объектов (П-О).

1. Для П-О 1:  $A_1^1 = \langle \text{ПСТМ имеют физико-механические свойства} \rangle;$

$A_2^1 = \langle \text{ПСТМ имеют свойство высокой износостойкости} \rangle;$

$A_3^1 = \langle \text{Физико-механические свойства ПСТМ могут обладать характеристикой изотропии} \rangle;$

$A_4^1 = \langle \text{ПСТМ обладают свойством ударной вязкости со значением, большим чем значение у монокристаллов} \rangle.$

2. Для П-О (категории) 2:  $A_1^2 = \langle \text{Температура может иметь неопределённое значение "высокая"} \rangle$ .
3. Для П-О (категории) 3:  $A_1^3 = \langle \text{Давление может иметь неопределённое значение "высокое"} \rangle$ ;  
 $A_2^3 = \langle \text{Давление может иметь неопределённое значение "высокое статическое"} \rangle$ ;  
 $A_3^3 = \langle \text{Давление может иметь неопределённое значение "низкое"} \rangle$ .

Аксиомы для понятий-процессов (П-П).

1. Для П-П 2:  $A_1^2 = \langle \text{Процесс получения ПСТМ I группы протекает при высокой температуре} \rangle$ ;  
 $A_2^2 = \langle \text{Процесс получения ПСТМ I группы протекает при высоком давлении} \rangle$ .
2. Для П-П 3:  $A_1^3 = \langle \text{Процесс получения ПСТМ II группы протекает при высокой температуре} \rangle$ ;  
 $A_2^3 = \langle \text{Процесс получения ПСТМ II группы протекает при высоком статическом давлении} \rangle$ .
3. Для П-П 4:  $A_1^4 = \langle \text{Процесс получения ПСТМ III группы протекает при высокой температуре} \rangle$ ;  
 $A_2^4 = \langle \text{Процесс получения ПСТМ III группы протекает при высоком статическом давлении} \rangle$ .
4. Для П-П 5:  $A_1^5 = \langle \text{Процесс получения ПСТМ IV группы протекает при низком давлении} \rangle$ .

Выделенные понятия-объекты, понятия-процессы, отношения  $\phi$  и  $R$  могут быть представлены как фрагмент онтологии, построенной на основе анализа ЕЯО. На рисунке 1 представлен фрагмент онтологии ПдО "Сверхтвёрдые материалы", а точнее – её раздела "Поликристаллические сверхтвёрдые материалы". Он построен автоматизированным способом на основе знаний, содержащихся в приведенном выше ЕЯО.

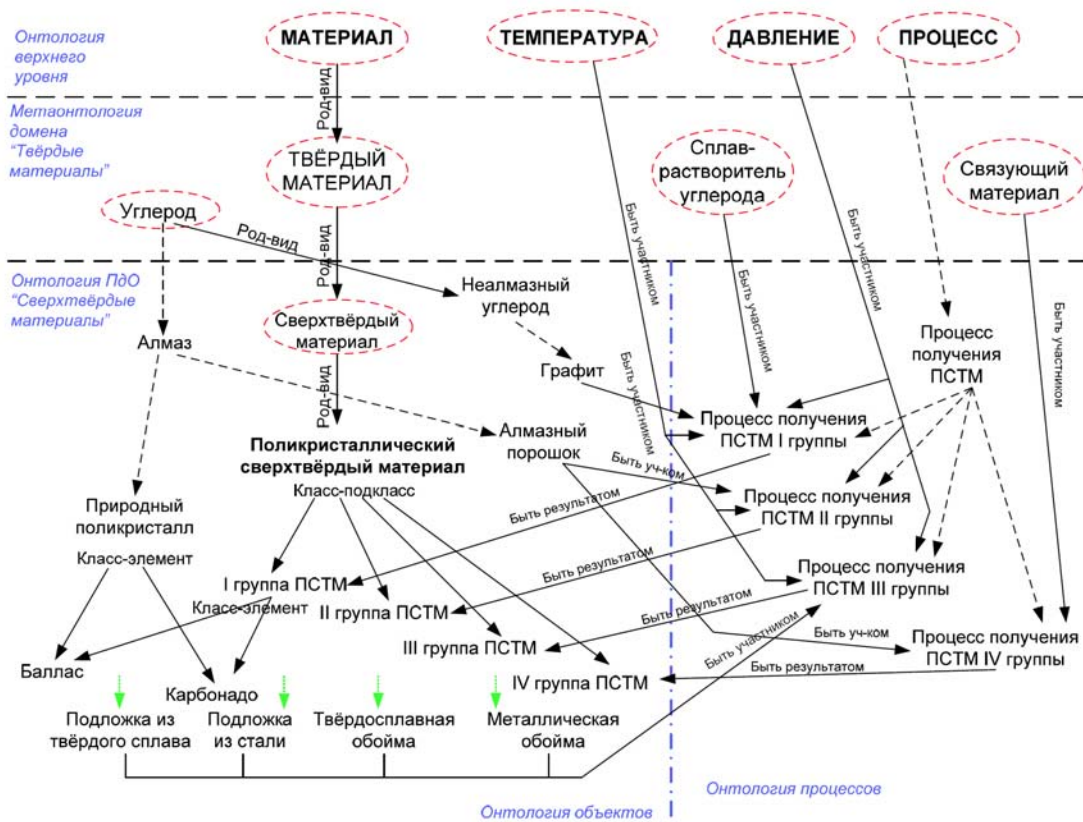


Рисунок 1. Фрагмент онтологии ПдО "Сверхтвёрдые материалы"

Процесс построения онтологии ПдО (в том числе и автоматизированным способом) предполагает этап начального развития онтологии. На этом этапе инженер по знаниям, совместно с экспертом в данной

ПдО, определяет список категорий онтологии верхнего уровня и метаонтологии домена "Твёрдые материалы", которые непосредственно связаны с концептами ПдО отношениями "выше-ниже". Далее эти категории структурируются и составляют начальное развитие (совместно с базовыми концептами ПдО) онтологического графа (ОГ). На рисунке вершины ОГ, соответствующие категориям, заключены в эллипсы с пунктирными линиями. Ребра ОГ, выполненные пунктирными линиями, в фрагменте текста в явном виде не указаны и вводятся как наиболее вероятные инженером по знаниям. При последующей обработке других текстов их следует уточнить (подтвердить или опровергнуть). На Рис.1 ОГ ПдО разделён на онтологии объектов и процессов штрихпунктирной линией.

---

## Выводы

---

Данная работа посвящена проектированию автоматизированных систем приобретения и обработки знаний с применением онтологий предметных областей. Полученные в результате автоматического анализа текста множества предметных констант, построенные отношения-предикаты являются исходными данными для построения онтологий предметной области. Формальное представление ЕЯО позволяет облегчить работу эксперта по составлению или уточнению онтологии. Дальнейшее направление исследований связано с уточнением семантики языка логического типа для представления знаний с целью выполнения различных операций логического вывода (получения новых знаний) на основе анализа ЕЯО и использованием в определении семантических отношений результатов работы системы анализа ЕЯО на основе вхождения слов[8].

---

## Благодарности

---

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теории и Приложений FOI ITHEA и Консорциума FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

## Литература

---

1. Кулик Б.А. Логика естественных рассуждений. С.-П.: Невский диалект.-2001.-127 с.
2. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах.-М.:Наука.-1989.-191с.
3. Тейз А., Грибомон П., Юлен Г. И др. Логический подход к искусственному интеллекту: От модальной логики к логике баз данных.-М.: Мир.-1998.-492с.
4. Широков В.А. Феноменологія лексикографічних систем. К.:Наукова думка.-2004.-327с.
5. Апресян Ю.Д. и др. Лингвистический процессор для сложных информационных систем.-М.: Наука.-1992.-287с.
6. Новые материалы. Колл. авторов. Под научной редакцией Ю.С. Карабасова. – М.: МИСИС, 2002 - 736 с.
7. Палагін О.В., Світла С.Ю., Петренко М.Г., Величко В.Ю. Про один підхід до аналізу та розуміння природномовних об'єктів. Комп'ютерні засоби, мережі та системи. -2008, №7. с.128-137.
8. Палагин А.В., Кривый С.Л., Бибилов Д.С. Обработка предложений естественного языка с использованием словарей и частоты появления слов 10 стр. (в печати).

---

## Информация об авторах

---

**Палагин Александр Васильевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [palagin\\_a@ukr.net](mailto:palagin_a@ukr.net)

**Кривый Сергей Лукьянович** – Киевский национальный университет им. Т.Г. Шевченко, Киев, ГСП, 01601, ул. Владимирская, 64, e-mail: [krivoi@i.com.ua](mailto:krivoi@i.com.ua)

**Величко Виталий Юрьевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [vitaly@aduis.kiev.ua](mailto:vitaly@aduis.kiev.ua)

**Петренко Николай Григорьевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [petrnq@ukr.net](mailto:petrnq@ukr.net)

## ОБРАБОТКА ПРЕДЛОЖЕНИЙ ЕСТЕСТВЕННОГО ЯЗЫКА С ИСПОЛЬЗОВАНИЕМ СЛОВАРЕЙ И ЧАСТОТЫ ПОЯВЛЕНИЯ СЛОВ

**Александр Палагин, Сергей Кривый, Дмитрий Бибииков**

**Аннотация:** *Описывается один из подходов к анализу естественно-языкового текста, который использует толковый словарь естественного языка, локальный словарь анализируемого текста и частотные характеристики слов в этом тексте.*

**Ключевые слова:** *представление текста, обработка текста, формальная логическая система.*

**ACM Classification Keywords:** *1.2.4 Knowledge Representation Formalisms and Methods - Representation languages, 1.2.7 Natural Language Processing - Language models*

**Conference:** *The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009*

---

### Введение

---

Проблемы, связанные с анализом естественно-языковых объектов (ЕЯО) традиционно относят к области искусственного интеллекта. Однако, объективные трудности, возникающие на пути анализа ЕЯО, не позволяют удовлетворительно решать проблему автоматизации такого анализа. Эти трудности связаны с тем, что проблема анализа ЕЯО относится к проблемам, которые плохо поддаются формализации. По проблеме анализа ЕЯО существует огромная литература, в которой описываются различные методы и подходы к решению частных случаев данной проблемы [см. 1-5]. В данной работе решается задача семантического анализа предложений естественного языка с целью извлечения знания, имеющегося в анализируемом тексте. При этом семантический анализ ограничивается одним предложением, и не рассматривается вопрос связей между предложениями, хотя некоторые предположения на этот счет делаются.

---

### 1. Общая постановка задачи анализа ЕЯО

---

Описываемая ниже общая постановка задачи анализа ЕЯО была сформулирована в работе [6] и здесь она конкретизируется для решения анализа предложений естественного языка.

Пусть  $T = t_1 t_2 \dots t_n$  естественно-языковой текст в алфавите  $X$ , т.е.  $T \in L(X)$ , где  $L(X)$  - язык над алфавитом  $X$ , а  $t_i \in T$  - предложения,  $i = 1, 2, \dots, n$ .

Каждое предложение  $t_i \in T$ , в свою очередь, имеет структуру  $t_i = t_{i_1} t_{i_2} \dots t_{i_m}$ , где  $t_{i_j}$  содержательно означают грамматические единицы, из которых построено предложение  $t_i$ . Если  $t_{i_j} \in t_i$ , то  $C_L(t_{i_j}) = t_{i_1} \dots t_{i_{j-1}}$  и  $C_R(t_{i_j}) = t_{i_{j+1}} \dots t_{i_m}$  будем называть левым и правым контекстом слова  $t_{i_j}$  соответственно в предложении  $t_i$ .

С текстом  $T$  свяжем такие объекты:

- $S$  - словарь языка  $L(X)$ , где содержатся слова  $t_{i_j}$  со своими определениями;
- $\gamma \subseteq T \times S$  - отношение, определяющее возможные значения и типы слова в словаре  $S$ ;

- $A = (D, \Pi)$  - предметная модель, на которой интерпретируется текст  $T$  ;
- $\phi \subseteq T \times A$  - отношение интерпретации текста  $T$  на модели  $A = (D, \Pi)$  .

Сигнатура предикатов  $\Pi = \{\pi_{k_1}, \dots, \pi_{k_r}\}$  содержит атомарные предикаты, из которых можно строить сложные формулы. Сейчас мы не будем фиксировать эту сигнатуру, поскольку она зависит от предметной модели. В связи с тем, что модель не уточняется, то и её сигнатуру уточнить нельзя. Заметим только то, что каждый атомарный предикат имеет тип (т.е. это будет некоторая типизированная сигнатура).

Определим теперь правила вычисления отношений  $\gamma$  и  $\phi$  .

Отношение  $\gamma$  имеет достаточно простой способ вычисления:

$$\gamma(t_{i_j}) = \{(a_1, \tau_1), (a_2, \tau_2), \dots, (a_s, \tau_s)\},$$

где  $a_i$  возможные значения слова  $t_{i_j}$  , а  $\tau_i$  - его возможные типы. Может случиться, что  $\gamma(t_{i_j}) = \emptyset$  . В этом случае значение этого слова считается неопределенным (и это требует пополнения словаря  $S$  ).

Отношение  $\phi$  определяется несколько сложнее. Если модель  $A = (D, \Pi)$  определена, то

$$\phi(T) = \phi(t_1) \dots \phi(t_n), \text{ где}$$

$$\phi(t_i) = \{\phi(\gamma(t_i)\gamma(C_R(t_i))), \phi(\gamma(C_R(t_i))\gamma(t_i)\gamma(C_R(t_i))), \dots, \phi(\gamma(C_L(t_i))\gamma(t_i))\},$$

при этом

$$\phi(\gamma(t_{i_j})) = \gamma(\phi(t_{i_j}));$$

$$\phi(\gamma(C_L(t_{i_j}))) = C_L(\gamma(\phi(t_{i_j})));$$

$$\phi(\gamma(\pi_r^k(p_1, \dots, p_k))) = \gamma(\phi(\pi_r^k))(\phi(\gamma(p_1), \dots, \gamma(p_k))),$$

где  $\gamma(\phi(\pi_r^k))$  - имя предиката, тип которого согласован с аргументами  $\gamma(p_1), \dots, \gamma(p_k)$  .

Из этой формальной постановки проблемы анализа ЕЯО вытекает, что основные задачи сводятся к таким:

- построить предметную модель  $A$  ; эта задача является основной и наиболее трудной в связи с тем, что предметная модель является по существу базой знаний (построение такой базы состоит в том, чтобы определиться с объектами, которые извлекаются из текста, с формальным логическим языком, правилами вывода, аксиоматикой и пр.);
- показать вычислимость отношений  $\gamma$  и  $\phi$  на предметной модели  $A$  и построить алгоритмы вычисления отношений  $\gamma$  и  $\phi$  ;
- при вычислении отношений  $\gamma$  и  $\phi$  контролировать соответствие типов аргументов и предикатов;
- определить взаимодействие алгоритмов вычисления  $\gamma$  и  $\phi$  с системами синтаксического и семантического анализа текста.

Второстепенными, но тоже важными, являются задачи связанные:

- с определением структуры данных для словарей;
- с определением информации, которая должна содержаться в словарях;
- с определением режима взаимодействия с пользователем (автоматический, полуавтоматический, диалоговый);
- язык пользовательского интерфейса и алгоритмы логического вывода.

## 2. Конкретизация общей задачи и схема ее решения

Нормализация ЕЯО. Очевидно, что в такой постановке проблема анализа ЕЯО носит очень общий характер, поэтому задачу необходимо конкретизировать. Рассмотрим два подхода к такой конкретизации. Большинство систем как информационного поиска, так и обработки текстовой информации содержат в качестве основного компонента систему анализа, служащую для выявления «содержания» или «значения» заданной единицы информации. В обычных системах такого рода анализ может выполнять человек. При этом он использует заранее разработанные таблицы или шаблоны для определения того, какой идентификатор содержания больше подходит по смыслу для заданной единицы информации. Известны также системы так называемого автоматического индексирования, в которых идентификаторы содержания приписываются автоматически, исходя из структуры текста документа и запроса.

В связи с тем, что естественный язык содержит различного рода нерегулярные явления, которые наблюдаются как в синтаксисе, так и в семантике, то система смыслового анализа должна приводить входные тексты к некоторому нормализованному виду, преобразуя различные, возможно неоднозначные, структуры на входе в фиксированные, стандартные идентификаторы содержания. Такого рода процедуры нормализации языка часто используют словари и списки слов, содержащие допустимые идентификаторы содержания, причем для каждого идентификатора приводится соответствующее определение с тем, чтобы регулировать и контролировать его использование. Следует заметить, что до появления понятия «онтология» процедуры анализа ЕЯО редко выходили за рамки анализа одного предложения. Это объясняется тем, что проблема анализа ЕЯО очень сложна и приходится сильно ограничивать свои запросы при попытке автоматизации такого рода анализа, выполняя некоторый упрощенный анализ текста. Рассмотрим такое упрощение, описываемое в данной работе.

Конкретизация задачи анализа. Конкретизации задачи анализа в нашем случае сводится к следующему.

Словарь  $S$ , о котором говорилось выше, является толковым словарем языка  $L(X)$  (это может быть словарь русского, украинского, английского или какого-либо другого естественного языка).

Текст  $T$  состоит из предложений языка  $L(X)$  и представляет текст, который не содержит никаких символов, кроме символов алфавита  $X$  (т.е.  $T$  не содержит формул, графиков, рисунков и т.п.).

Отношение  $\gamma$  состоит из суперпозиции двух отношений  $\gamma_1 * \gamma_2$ , выполняемых последовательно. Содержательно отношение  $\gamma_1$  означает распознавание принадлежности слова к данному языку и проверку правильности написания слова  $t_{i_j} \in t_i$ , где  $t_i \in T$  в соответствии с написанием его в толковом словаре, т.е.

$$\gamma_1(t_{i_j}) = \begin{cases} 1, & \text{если } t_{i_j} \in S; \\ 0, & \text{если } t_{i_j} \notin S. \end{cases}$$

Если слово  $t_{i_j} \in t_i$  распознано в словаре  $S$ , то оно заносится в словарь  $T'$  правильных слов, а если это не так, то предусматривается сигнализация о том, что данное слово отсутствует в словаре  $S$  и принимается решение о добавлении данного слова в словарь или его исправлении (слово может быть искажено, например, вследствие сканирования текста  $T$ ).

Словари  $S$  и  $T'$  являются входными данными для отношения  $\gamma_2$ . Содержательный смысл отношения  $\gamma_2$  сводится к тому, что если  $\gamma_1(t_{i_j}) = 1$ , то  $\gamma_2(t_{i_j})$  определяет его грамматическую единицу языка (имя

собственное, сказуемое, существительное, числительное и т.п.) а также возможные флексии слова  $t_{i_j} \in t_i$ .

Областью интерпретации текста  $T$  является модель  $A = (D, \Pi)$ , где  $T$  - это исходный текст, возможно расширенный некоторой дополнительной информацией, а сигнатура  $\Pi$  определяется из текста  $T$  в результате использования информации о различных вхождениях слова  $t_{i_j}$  в предложения  $t_i \in T$ . При этом вычисление отношения  $\phi$  ограничивается отдельно взятым предложением  $t_i \in T$ , определяемым каждым вхождением слова  $t_{i_j}$  в текст  $T$ . В случае трудности определения предиката  $\pi_i \in \Pi$ , предусматривается диалоговый режим вычисления  $\phi(\pi_i)$  и  $\gamma(\phi(\pi_i))$ .

Схематично предлагаемая система анализа выглядит следующим образом:

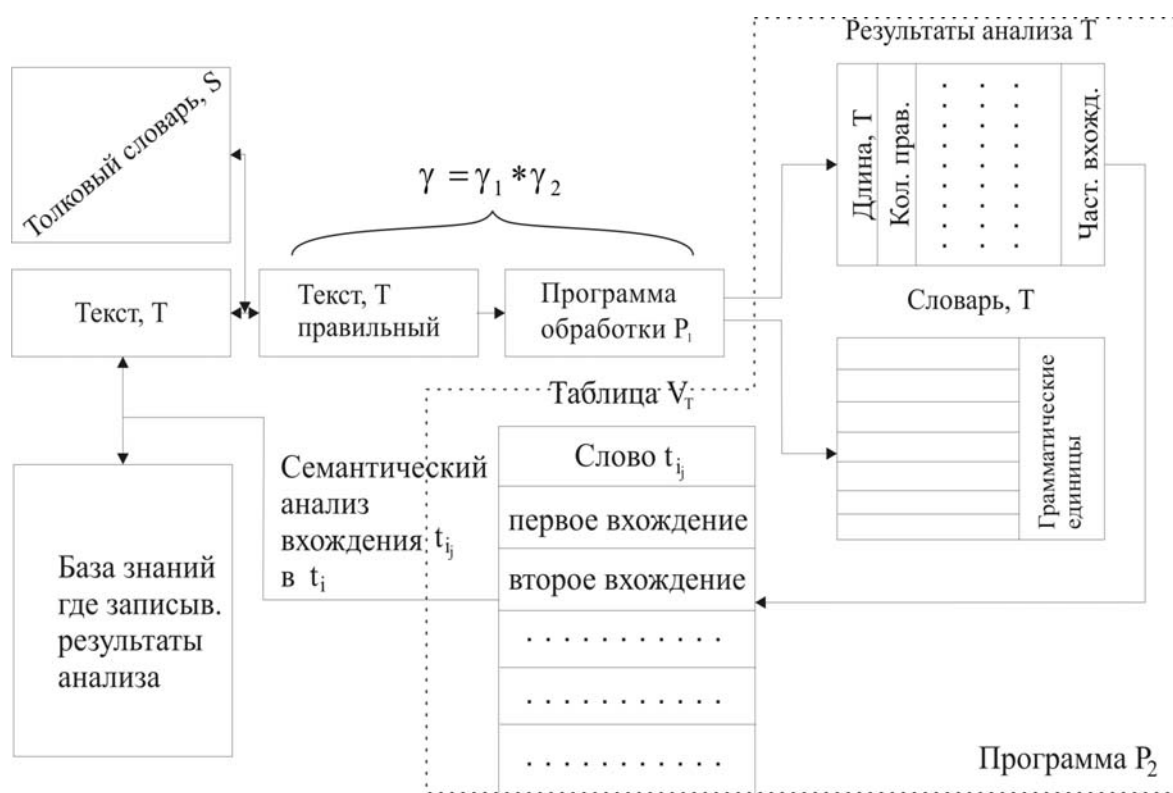


Рис.1. Схема системы анализа предложений естеств. языка

В этой схеме отношение  $\gamma = \gamma_1 * \gamma_2$  вычисляет программа  $P_1$ . Результатом ее работы является два файла  $F_1$  и  $F_2$ , заполненные соответственно числовыми характеристиками слов входного текста  $T$  и словами  $t_{i_j}$  предложений этого текста. Структура файла  $F_2$  показана ниже на рисунке 2.



Слова	Длина	Частота	?Ч
общественно-политической	24	1	Распознано
специализацииивысшим	24	1	Распознано
социально-экономического	24	2	Распознано
социально-экономических	23	2	Распознано
национально-культурного	23	1	Распознано
ответственностираздел	22	1	Распознано
предпринимательская	21	1	Распознано
самоуправлениястатья	21	1	Распознано
председательствующего	21	1	Распознано
оперативно-розыскную	20	1	Распознано
нормативно-правовыми	20	1	Распознано
общегосударственными	20	1	Распознано
нормативно-правовые	19	1	Распознано
предпринимательскую	19	1	Распознано
предпринимательской	19	2	Распознано
общегосударственных	19	2	Распознано
научно-технического	19	2	Распознано
предпринимательства	19	1	Распознано
внешнеэкономической	19	2	Распознано
общегосударственные	19	1	Распознано
нормативно-правовым	19	1	Распознано

Рис. 2. Структура файла  $F_2$ 

Файлы  $F_1$  и  $F_2$ , сформированные программой  $P_1$ , служат входными данными для работы программы  $P_2$ , которая вычисляет отношение  $\phi$ . При этом, работа программы  $P_2$  сводится к построению таблицы  $V_T$  для слов  $t_{ij} \in t_i$ ,  $t_i \in T$ . Затем, по этой таблице и предложениям текста  $T$  определяется семантический смысл рассматриваемого предложения. Предложение  $t_i \in T$  определяется на основе номера вхождения слова  $t_{ij}$  в текст  $T$  с помощью таблицы  $V_T$ , вид которой приведен ниже на рисунке 3..

Слово	№ вхождения	№ предложения, $t_i$
Палагин	1	1
	2	2
	3	3
	4	5
Кривой	1	1
	2	4
Петренко	1	1
	2	2
Яковлев	1	3
Опанасенко	1	5
Кургаев	1	6

Рис. 3. Структура таблицы  $V_T$

### 3. Пример работы системы

Рассмотрим в качестве примера текста  $T$  текст некоторой библиографической информации и анализ этого текста с помощью описанной системы (например, тексты статей, присылаемых на конференцию, и их библиография). Такого рода текст имеет четко нормализованную структуру: нулевая позиция – универсальный определитель (УДК), первая часть текста – список авторов, вторая часть – название статьи, третья часть – издательство, четвертая – год издания.

Тогда построение модели  $A = (D, \Pi)$  сводится к наполнению данными множеств УДК, АВТОР, СТАТЬЯ или КНИГА, ИЗД-ВО, ГОД, РЕЦ, где множество РЕЦ означает множество рецензентов, которое предположительно имеется и построено пользователем системы. Это наполнение в данном случае выполняется автоматически, с использованием имеющейся таблицы  $V_T$ . Более того, использование информации о вхождении того или иного слова в текст  $T$  позволяет автоматически построить декартово произведение

$$D = \text{УДК} \times \text{АВТОР} \times (\text{СТАТЬЯ} \cup \text{КНИГА}) \times \text{ИЗД-ВО} \times \text{ГОД} \times \text{РЕЦ}.$$

Построив это множество, можно строить сигнатуру предикатов, хотя в данном случае в этом нет необходимости. Это объясняется тем, что из множества  $D$  средствами реляционной базы знаний можно построить отношения, которые нас интересуют. Например, используя селекцию проекции множества  $D$  по атрибутам АВТОР, СТАТЬЯ, РЕЦ получаем отношение-предикат  $R_{acp}$ , содержащее информацию о статьях, авторах и рецензентах статей этих авторов. Приведем конкретные значения текста и результатов его анализа.

Пусть исходный текст представляет список литературы и список РЕЦензентов, которые рецензировали эти работы:

1. Палагин А.В., Кривой С.Л., Петренко Н.Г. статья1. - изд1. - 2009. - С.1-10. - (рец. P1).
2. Палагин А.В., Петренко Н.Г. статья2. - изд2. - 2007. - С. 21-30. - (рец. P2).
3. Палагин А.В., Яковлев Ю.С. книга1. - изд3. - 2002. - 500 с. - (рец. P3).
4. Кривой С.Л. статья3. - изд3. - 2007. - С. 5-17. - (рец. P4).
5. Палагин А.В., Опанасенко В.Н. книга2. - изд4. - 2004. – 300 с. - (рец. P2).
6. Кургаев А.Ф. книга 3. - изд5. - 2007. - 450 с. - (рец. P3).

Анализ этого текста дает такие множества предметных констант и их семантические значения:

$$\begin{aligned} \text{АВТОР} &= \{\text{Палагин А.В., Кривой С.Л., Петренко Н.Г., Яковлев Ю.С., Опанасенко В.Н., Кургаев А.Ф.}\}, \\ \text{СТАТЬЯ} \cup \text{КНИГИ} &= \{\text{статья1, статья2, статья3, книга1, книга2, книга3}\}, \\ \text{ИЗД} &= \{\text{изд1, изд2, изд3, изд4, изд5}\}, \\ \text{РЕЦ} &= \{P1, P2, P3, P4\}. \end{aligned}$$

Применяя оператор декартового произведения отношений реляционной алгебры, получаем множество  $D$ . Из этого множества тем же способом с помощью операторов проекции и селекции получаем отношения:

$$R_a = \{\text{Палагин А.В., Кривой С.Л., Петренко Н.Г., Яковлев Ю.С., Опанасенко В.Н., Кургаев А.Ф.}\},$$

элементы которого упорядочены так, как они идут в этом отношении. Возможная интерпретация констант в словаре  $S$  может выглядеть так:

$$\gamma (\text{автор1}) = (\text{Палагин А.В., имя-собственное, дтн, профессор, академик,...})$$

Аналогично по остальным авторам.

Далее из множества  $D$  определяются отношения-предикаты: отношение "автор-статья"

$$R_{ac} \subseteq \text{АВТОР} \times (\text{СТАТЬЯ} \cup \text{КНИГИ}),$$

семантическое значение которого имеет вид:

$$R_{ac} = \{(\text{Палагин А.В., статья1}), (\text{Палагин А.В., статья2}), (\text{Палагин А.В., книга1}),$$

$$(\text{Палагин А.В., книга2}), (\text{Кривой С.Л., статья1}), (\text{Кривой С.Л., статья3}),$$

$$(\text{Петренко Н.Г., статья1}), (\text{Петренко Н.Г., статья2}), (\text{Яковлев Ю.С., книга1}),$$

$$(\text{Опанасенко В.Н., книга2}), (\text{Кургаев А.Ф., книга3})\};$$

отношение "автор-рецензент-статья"

$$R_{аср} \subseteq \text{АВТОР} \times \text{РЕЦ} \times \text{СТАТЬИ} \cup \text{АВТОР} \times \text{РЕЦ} \times \text{КНИГИ},$$

семантическое значение которого имеет вид:

$$R_{аср} = \{$$

(Палагин А.В., Р2, статья1), (Палагин А.В., Р1, статья2),

(Палагин А.В., Р2, книга1), (Палагин А.В., Р3, книга2),

(Палагин А.В., Р3, книга1), (Палагин А.В., Р1, книга2),

(Палагин А.В., Р4, книга1), (Палагин А.В., Р2, книга2),

(Кривой С.Л., Р4, статья1), (Кривой С.Л., Р4, статья3),

(Петренко Н.Г., Р2, статья1), (Петренко Н.Г., Р3, статья2),

(Яковлев Ю.С., Р2, книга1), (Опанасенко В.Н., Р3, книга2),

(Яковлев Ю.С., Р3, книга1), (Опанасенко В.Н., Р4, книга2),

(Яковлев Ю.С., Р4, книга1), (Опанасенко В.Н., Р1, книга2),

(Кургаев А.Ф., Р1, книга3), (Кургаев А.Ф., Р2, книга3),

(Кургаев А.Ф., Р4, книга3)\}.

Первое отношение бинарное, а второе - тернарное. Аналогичным способом можно получить и другие интересующие пользователя отношения-предикаты, т.е. полностью построить сигнатуру предикатов  $\Pi$ , исходя из области  $D$ .

---

## Заключение

---

Данная работа посвящена проектированию автоматизированных систем извлечения и обработки знаний, а также онтологий предметных областей. Она является продолжением работ [6-8], где описывались технические возможности реализации такого типа систем.

Какие задачи можно решать с помощью предлагаемой системы? Поскольку данная система является только первым этапом на пути создания автоматизированных систем извлечения и обработки знаний, то эта система извлекает информацию из ЕЯО, необходимую для построения онтологий предметной области, к которой относится анализируемый текст. А именно,

- а) распознается смысл данного конкретного слова в зависимости от того, в какое предложение оно входит;
- б) строится модель  $A = (D, \Pi)$ , элементы которой разбиваются по типам (концепты, предикаты, константы и т. п.). После такого разбиения эта модель является основой для построения онтологии данной предметной области или пополнения уже существующих онтологий;
- в) позволяет ввести время, в котором происходят события в данном тексте.

Следующим шагом на пути анализа ЕЯО является использование определений с целью построения онтологии предметной области или пополнения уже существующей. Эта идея была предложена в работе [9], которая базировалась на использовании теории конечных автоматов. Пополнение имеющейся онтологии связывалось с операцией итерации, а соединение нескольких онтологий в одну (интеграция онтологий) – с операциями объединения, пересечения и конкатенации автоматов, представляющих данные онтологии. Для пояснения этого подхода рассмотрим пример.

Пусть строится онтология «Транспортные средства». Известно, что транспортные средства делятся на наземные, морские и воздушные. В свою очередь наземные транспортные средства делятся на рельсовые, колёсные (безрельсовые) и гужевые. К рельсовым транспортным средствам относятся поезд, трамвай, к безрельсовым – автомобиль, автобус, троллейбус, к гужевым – верблюд, лошадь. Эта классификация дает возможность построить такую онтологию:



Пусть в тексте встречаются определения:

**Тролейбус – безрельсовое транспортное средство, приводимое в движение электродвигателем.**

**Автобус – колесное транспортное средство, приводимое в движение двигателем внутреннего сгорания.**

Тогда система анализа строит такую конструкцию:

$$\text{Тролейбус}(x) \Leftrightarrow \text{Трансп.средство}(x) \wedge \neg(\text{Рельсовое}) \wedge \exists y(\text{Прив} - \text{в} - \text{Движ}(x, y) \wedge \text{Электродв}(y))$$

По этой конструкции система находит в имеющейся онтологии концепты «Трансп. средство», «(безрельсовое) колесное» и «тролейбус». Отсутствие в данной онтологии концепта «электродвигатель» приводит к тому, что система ищет онтологию «Двигатели» или «Электродвигатели». Если находит, то строится предикат  $\exists y(\text{Прив} - \text{в} - \text{Движ}(x, y))$ , где  $x$  и  $y$  конкретизированные концепты-переменные. В противном случае система сигнализирует о неполноте существующей онтологии, которую необходимо пополнить понятием «Двигатели». После пополнения имеющейся онтологии, анализ второго определения уже не приводит к ситуации неполноты имеющейся онтологии.



Заметим, что система анализируя определение, естественным образом находит иерархию понятий. Например, в определении троллейбус понятие «Трансп.средство» является более общим, чем понятие «безрельсовое». Отсюда следует, что нормализация текста должна сводиться к тому, что в определениях отношение следования необходимо задавать от «более общего» к «более конкретному».

---

## Благодарности

---

Работа выполнена при финансовой поддержке Министерства образования и науки Украины в рамках совместного Украинско-Болгарского проекта № **145 / 23.02.2009** «Разработка распределенных виртуальных лабораторий на основе прогрессивных методов доступа для поддержки проектирования сенсорных систем» и Болгарского национального научного фонда в рамках совместного Болгарско-Украинского проекта **D 002-331 / 19.12.2008** с тем же названием.

---

## Литература

---

1. Кулик Б.А. Логика естественных рассуждений. С.-П.: Невский диалект.-2001.-127 с.
2. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах.-М.:Наука.-1989.-191с.
3. Тейз А., Грибомон П., Юлен Г. И др. Логический подход к искусственному интеллекту: От модальной логики к логике баз данных.-М.: Мир.-1998.-492с.
4. Широков В.А. Феноменологія лексикографічних систем. К.:Наукова думка.-2004.-327с.
5. Апресян Ю.Д. и др. Лингвистический процессор для сложных информационных систем.-М.: Наука.-1992.-287с.
6. Палагин А.В., Крывый С.Л., Петренко Н.Г. Знаниеориентированные информационные системы обработки знаний, содержащихся в естественно-языковых объектах: методологические основы и архитектурная организация. – УсиМ.-2009. (в печати)
7. Палагин А.В. Архітектура онтологоуправляемых компьютерных систем. - Киберн. и сист. анализ.- 2006.-№2.- С.111-124.
8. Палагін О.В., Петренко М.Г. Архітектурно-онтологічні принципи побудови інтелектуальних систем. - Математичні машини і системи.-2006.-№4.- С.15-20.
9. Крывый С.Л., Ходзинский А.Н. Автоматное представление онтологий и операции на онтологиях. –International Book Series. – N 1. – Algorithmic and Mathematical Foundations of the Artificial Inetelligence. – ITHEA: Sofia. -2008. – PP. 173-179.

---

## Информация об авторах

---

**Палагин Александр Васильевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [palagin\\_a@ukr.net](mailto:palagin_a@ukr.net)

**Крывый Сергей Лукьянович** – Киевский национальный университет им. Т.Г. Шевченко, Киев, ГСП, 01601, ул. Владимирская, 64, e-mail: [krivoi@i.com.ua](mailto:krivoi@i.com.ua)

**Бибиков Дмитрий Сергеевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [bb\\_coff@mail.ru](mailto:bb_coff@mail.ru)

## ОПРЕДЕЛЕНИЕ ПОНЯТИЯ «СМЫСЛ» ЧЕРЕЗ ОНТОЛОГИЮ. СЕМАНТИЧЕСКИЙ АНАЛИЗ ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА

Леонид Святогор, Виктор Гладун

**Аннотация:** Предложен новый подход к понятию «смысл» и дано его формально-графическое определение через онтологию. Рассмотрена задача семантического (смыслового) анализа текстов ЕЯ, которая основана на процедуре поиске подграфа концептуального графа, отображающего знания о мире.

**Ключевые слова:** онтология, смысл, текст, семантический анализ.

**ACM Classification Keywords:** 1.2.4 Knowledge Representation Formalisms and Methods

**Conference:** The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009

---

### Введение

Практически востребованным в проблеме искусственного интеллекта и в создании интеллектуальных компьютеров является решение задачи семантического анализа естественно-языковых (ЕЯ) текстов. В известных информационных системах эту задачу рассматривают на уровне архитектуры лингвистического процессора, выполняющего синтаксический, морфологический и семантический анализ с использованием онтологии [1]. Предложены также иные схемы, в которых анализ текста основывается на использовании систем репрезентации общих или профессиональных знаний в виде онтологий [2].

Развитие в этом направлении наталкивается на основную трудность: что нужно считать результатом семантического анализа текста? В данной работе поставлена задача – определить результат анализа текста через категорию «смысл». Для этого, во-первых, интуитивное понимание «смысла сообщения», или его «содержания», заменяется формальным определением данного понятия. Во-вторых, предлагается конструктивная процедура обнаружения и лексикографического отображения смыслового содержания текстов, предъявленных для семантического анализа.

---

### Предпосылки для пересмотра устоявшихся представлений

Интуитивно мы понимаем смысл как меру «благоразумного» поведения интеллекта, адекватную сложившейся ситуации, а при коммуникации – как адекватное отображение реальности. Для создания машинного интеллекта необходимо решить амбициозную задачу: определить смысл как формальную категорию, не удаляясь от её привычного человеческого понимания.

Благодаря усилиям психологов, а также лингвистов, решающих практическую задачу машинного перевода, были выдвинуты две философские концепции: (а) функция естественного языка заключается в «содержательной организации и представлении знаний» (Городецкий Б.Ю., 1989) и (б) целью коммуникативного акта является передача замысла от автора к потребителю [6]. Не касаясь трудностей общения, связанных с согласованием баз знаний коммуникантов и их выразительных способностей, центральной проблемой назовём одну: как сформулировать искомый параметр сообщения – его «смысл», чтобы его можно было обнаружить в сообщении автоматически, без подсказок автора.

---

С появлением **теории предикатов**, математической лингвистики и контекстно-свободных грамматик Н. Хомского надежды на решение данной проблемы окрепли. Появилась уверенность, что для представления естественно-языковых объектов *«наиболее приемлемым на сегодняшний день является язык логического типа, в котором понятие связь по смыслу формализуется с максимальной полнотой»* (Кривой С.Л., 2009). И всё же отмечается, что *«в области логического анализа эмпирического знания успехи логики более чем скромны»*. На наш взгляд, это объясняется тем, что формально-грамматические конструкции решают чисто математическую задачу – выводимости и непротиворечивости слов заданного алфавита с правилами вывода и системой аксиомой. По определению, формальные языки являются бесконтекстными, то есть – не адекватными живому языку. Определение «реального смысла» переносится в иную область рассмотрения.

**Классическая лингвистика** выдвинула на первый план лингвистическую модель «Смысл–Текст» (Мельчук И.А., 1974). Кроме синтаксического, морфологического и фонологического компонентов в модели присутствует семантический компонент, в котором понятие «смысл» представлено графом предикатных отношений. В модели «Смысл–Текст» смысл определяется как инвариант синонимических преобразований отдельной фразы. ([3], стр. 10). Однако практическая реализация определения, особенно при переходе от фразы к тексту, реально невозможна. Критический анализ лингвистической модели «Смысл–Текст» приведен ниже.

В последнее время в теории искусственного интеллекта получили развитие **системы представления знаний** (*knowledge representation*) в виде семантических сетей – онтологий. Онтология представляет собой такую сеть, в которой вершинами являются понятия, выраженные лексическими средствами, а отношения отражают связи теоретико-множественного или неформально-семантического характера. Онтологии являются мощным инструментом описания внешнего и ментального мира, в который погружена деятельность человека. В принципе, эта деятельность является осмысленной, то есть – целенаправленной. Поэтому может быть выдвинута гипотеза о тесной связи знаний со смыслом. Эта гипотеза положена в основу данного исследования.

---

### Цель работы

---

До сих пор разработчики онтологий не ставили перед собой задачу связать знания со смыслом. Поэтому неизвестны попытки дать такое определение понятия «смысл», которое было бы конструктивным при анализе текстов дискурсивного характера. В работе предлагается концептуально новый подход к определению понятия «смысл».

**Целью настоящей работы** является найти такое формальное представление «смысла», которое удовлетворяло бы следующим условиям:

1. Смысл должен выражать в сжатом виде некоторое **знание**, т. е. – опираться на **онтологический базис**.
2. Смысл должен быть **формализован**. Его извлечение из текста должно производиться **алгоритмически**.

---

### Почему лингвистическая модель «Смысл–Текст» непригодна для процедурного анализа текста

---

Ключевым понятием рассматриваемой лингвистической модели «Смысл–Текст» [3] является графическая структура, отображающая для заданного предложения его смысл. Граф наполняется словами, взятыми непосредственно из предложения, а слова связываются между собой с помощью специальных символов



нетерминального алфавита, которые названы «смысловыми атомами», или семами, Вершинами семантического графа служат семы двух типов: имена предикатов и имена объектов или классов объектов. Семы – это не наблюдаемые, а постулируемые сущности; их вводят для преобразований, сохраняющих смысл. Они составляют базовый алфавит семантического словаря модели.

Позиция автора модели заключается в том, что *«естественно стремиться к тому, чтобы базовый алфавит содержал средства, достаточные для эксплицитного выражения любых, сколь угодно тонких смысловых различий, какие только могут встретиться в тексте и какие мы пожелаем выразить»* ([3], стр. 64). Фактически это означает, что семантический словарь должен отражать самые широкие знания о внеязыковой действительности. В то же время модель не содержит никакого конструктивного задания этой действительности: она обращена непосредственно к тексту и к интуиции исследователя.

Для того, чтобы изобразить графически смысловое содержание фразы, требуется извлечь из семантического словаря такое количество «смысловых атомов», которое намного превышает собственное количество слов в заданном предложении. Из-за этого граф становится достаточно громоздким. Например, чтобы отобразить смысл фразы **«Косте удалось победить»**, потребовалось использовать такие предикатные и объектные семы: «иметь место», «каузировать», «цель», «зависеть», «неверно», «ресурсы», «имя» и некоторые другие – всего 15 лексем. Семантический граф приведенной выше фразы содержит 22 узла и толкуется ([3], стр. 99) следующим образом:

***Этот мужчина по имени Костя каузирует действовать (т.е. пускает в ход) Игрек, являющийся его ресурсами, с целью победить Икса в Игрек, и это – наряду с Зет, не зависящим от Кости – каузирует победу Кости, которая имеет место однократно до момента речи.***

В терминологии использованы «укрупнённые» семы. *«Если попытаться разложить каждый элемент этого семантического графа до сем, то он, вероятно, не уместится на печатной странице»* ([3], стр.64).

С точки зрения лингвиста указанная интерпретация предложения, вероятно, является необходимо полной. Однако стремление отразить на семантическом графе взаимно-противоречивые требования, а именно – совместить его предельно «тонкую» смысловую выразительность и максимальную инвариантность к синонимии текста – приводят к громоздким построениям, несовместимым с практикой. Тем более, что при анализе текста как связанной совокупности фраз потребуется каким-то образом соединять смыслы отдельных предложений, и тогда вся смысловая графовая конструкция «рухнет».

В этой связи уместно обратить внимание на существенное замечание Дж. Хокинса относительно процессов мышления: *«Вспоминание сохраняется в форме, охватывающей **существенные связи** между элементами целого, а не преходящие детали»* ([4], стр. 84).

Второе замечание относится к аксиоматической базе модели «Смысл–Текст». Содержательная сторона аксиом и соответствие семантического представления реальной действительности не рассматриваются. Семантический язык модели *«предназначен вовсе не для логического анализа записанной на нём информации, а для удобного представления содержания любых высказываний – тривиальных, неопределённых, противоречивых и бессмысленных»*. ([3], стр. 73). Но в таком случае остаётся неясным, как проверить «смысл формального смысла», то есть – какое решение потребителя будет правильным?

Наконец, принципиальный, на наш взгляд, недостаток лингвистической модели кроется в её отрыве от базы знаний. Формальный смысл фразы рассматривается не в контексте общечеловеческого знания, а потому выступает как чистая категория теоретической лингвистики.

Таким образом, в результате анализа лингвистической модели «Смысл–Текст» было выяснено, что графическое представление смысла является громоздким, особенно когда речь идёт о слитном тексте. Аксиоматическая база безразлична к знаниям, а семантический анализ текста выполняется в отрыве от реальной действительности и потому лишён контекста.

Альтернативным подходом к определению и нахождению смысла является онтологический подход.

---

### Онтология как инструмент концептуальной организации знаний

---

Информационная функция онтологии выражена Дж. Хокинсом: «*Единственный способ, которым человек может понять этот изменчивый мир, – найти инвариантную структуру для переменного потока информации*» ([4], стр.85).

Попытки привести в систему понятия о мире, которые «рассыпаны» в естественных языках, предпринимались давно и завершились созданием языковых тезаурусов. Обратная задача – понять, как человек структурирует свои знания о мире, привела к разработке лексических онтологий, освободив этот термин от его чисто философского содержания.

Формально лексическая онтология представляется семантической сетью, в которой: (а) вершинами служат понятия естественного языка, (б) связями являются допустимые в языке отношения и (в) элементы онтологии интерпретируются внутри естественного языка.

К онтологиям исследователи шли двумя встречными путями. Абстрактные онтологии (СУМО, Микрокосмос, Дж. Соуы и др.) оперируют ресурсами высокого уровня общности и на этом уровне остаются. Предметные онтологии, наоборот, стремятся описать конкретную профессиональную область (ПрО). Между общими и предметными представлениями возник «вакуум». Однако, поскольку знания представляют собой целостную систему мировоззрения, то размежевание может быть оправдано только целью изучения его по частям, но при этом обязательно в контексте общего знания.

Неизбежно возник вопрос: каким образом объединить все знания в рамках единой системы? Ответом на поставленный вопрос явилось создание концептуально-иерархической трёхуровневой онтологии, которая описана ниже.

---

### Иерархическая трёхуровневая онтология

---

**Иерархической трёхуровневой онтологией (ИО\*3)** назовём семантическую сеть, в которой **вершинами** служат лексические обозначения категорий и понятий естественного языка, а **связи** организованы по принципу семантической целесообразности [2]. Лексические единицы названы **концептами**. Каждый концепт обозначается именем и может быть **интерпретирован** на множестве словарных статей естественного языка, или определён экспертом. Существенно, что имя концепта в предложенной онтологии отображает в языке такое понятие, которое имеет содержательный смысл и участвует в формировании знаний и представлений человека о внешнем и внутреннем мире.

Концепты и связи устанавливает эксперт. В онтологии ассоциаций ИО\*3 связи могут быть как формальными («часть–целое», «элемент–множество», «является»), так и неформальными («имеет признаки», «связано ассоциативно» и др.). С логической точки зрения эти отношения являются двух- и многоместными предикатами. В то же время все эти типы связей можно нивелировать и рассматривать в онтологии только один **универсальный предикат** – «**связь (А, В)**», где **А** и **В** – лексические концепты. Универсальный предикат на графе ИО\*3 принимает только одно значение – «единица».

Концепты онтологии могут группироваться в кластеры понятий. Что касается **аксиомы**, то в ИО\*3 ей соответствует имя самой высшей категория знания – *Материя*.

Наиболее существенной стороной онтологии ИО\*З является её трёхуровневая иерархическая структура, которая отражает сложный процесс освоения, накопления и обобщения человеческих знаний. Она имеет пирамидальную архитектуру. На самом верхнем уровне пирамиды обозначены основания мировоззрения, на нижнем – конкретные знания предметных областей.

На верхнем уровне онтологии обозначены **общие ресурсы** знания, такие категории как *Материя, Энергия, Жизнь* и другие. В основу деления положена материалистическая парадигма академика В.И. Вернадского о делении материи на *Косное* и *Живое вещество* и о делении живого на *Биосферу* и *Ноосферу*. Здесь, благодаря факту понятийной непересекаемости естественно-научных категорий, который принят современной наукой, реализован принцип классификации понятий.

Нижний уровень, наоборот, используется для представления **профессиональных знаний**, оформленных в предметных областях. Узлы и связи отражают концептуальные понятия, ключевые слова и терминологию данной ПрО, которые специалист считает нужным внести в онтологию.

Естественно, что кроме общих и профессиональных знаний должен быть представлен и слой так называемого «промежуточного знания», связывающего верхний уровень с нижним. Поэтому в ИО\*З вводится средний уровень онтологии. Здесь формируется **концептуальная среда междисциплинарного общения**, то есть – такие понятия, которые широко употребляются в информационной и традиционной человеческой деятельности.

Например, один из вариантов концептуального раскрытия онтологического термина «**Человеческое общество**» на среднем уровне может быть таким (фрагментарно):

*Социум* = (Личность; Организация; Базис; Надстройка; ...);

*Личность* = (Права; Обязанности; Свобода; Работа; Отдых; ...);

*Организация* = (Институты; Структурирование; Функционирование; ...);

*Институты* = (Органы власти (высшие; судебные; исполнительные; региональные; ...);  
Органы управления (Прокуратура, Полиция, Администрация; ...); Здоровье (...); Образование (...); Отдых (...); ...);

*Структурирование* = (Базис; Надстройка; ...);

*Функционирование* = (Экономика; Политика; Бизнес; Законодательство; ...); и т. д.

На графе могут быть показаны перекрёстные связи между концептами одного уровня подчинённости.

По сути сказанного, предложенная Иерархическая трёхуровневая онтология является графической схемой для объединения понятий ЕЯ, представленных их именами. Она служит универсальным способом структурно-иерархической репрезентации знаний разного типа и разной степени общности, которые могут быть обозначены лексическими единицами – концептами и связаны между собой семантическими отношениями. Являясь декларативной структурой, трёхуровневая онтология выступает в качестве составляющей баз знаний и, по нашему мнению, решающим образом участвует в процессе обнаружения смысла сообщений.

Ниже будет показано, каким образом описанная онтология связана с искомым определением понятия «смысл».

---

### Что такое «смысл»? Мнения специалистов

---

Общего определения здесь нет, и вряд ли оно может быть получено. Смысл выступает, с одной стороны, как объективная (экспертная) оценка наблюдаемой ситуации, а с другой – как ментальная, ненаблюдаемая категория сознания.

---

С психологической точки зрения смысл – это категория для оценки поведения разумного существа. Осмысленной является ситуация, в которой объекты и связи отвечают реальной действительности. Осмысленным является целенаправленное поведение, адекватное ситуации. В коммуникативном психологическом акте – это категория такого понимания, которое может быть проверено [4]. В интеллектуальной конфликтной модели В.А. Лефевра ([5], 2003) смысл трактуется как разумное поведение личности в данной этической системе.

С лингвистической точки зрения смыслом является категория семантики языка, которая отвечает за согласованность лексики с внешним миром и отражёнными в языке знаниями, а также управляет прагматикой для приведения её в соответствие семантике. Нахождение смысла увязано с «распознаванием стратегического замысла говорящего» ([6], стр. 26).

В литературе по нейролингвистике рассматривают смысл в контексте процесса отображения слуховых и зрительных образов (внешних данных) в структуры памяти, локализованные или распределённые в мозге [7]. Ментальный смысл представлен «мыслительными единицами», которые получаются в результате преобразования языковых единиц и согласования их с внутренними представлениями (ментальными моделями знания). В числе мыслительных единиц выделяют **узуальный смысл**, который априори согласуется с внутренней картиной мира, и **окказиональный смысл**, который апостериорно ассоциируется с узуальным смыслом путём мысленного напряжения. Окказиональный смысл является реакцией индивидуума на сообщение; он либо извлекается из глубины памяти, либо закрепляется как новое знание.

С точки зрения нейрофизиологии объектом знания является «ситуация», которая может быть формально определена через «объекты», «атрибуты» и «действия». ([8], стр. 82) Образ внешней среды отображается (вербализуется) в структурное описание, которое задано совокупностью отдельных образных представлений элементарных ситуаций. Характерно, что отдельное образное представление представляет собой кластер «мыслительных единиц», который охватывает не более семи вершин. В такой интерпретации понятию «смысл», по-видимому, можно поставить в соответствие подструктуру мыслительных единиц – совокупность нейронных кластеров, связанных между собой активными связями. Такие подструктуры носят ментальный характер, но способны генерировать текстовое описание ситуации.

---

### Что такое онтологический смысл?

---

На фоне указанных выше представлений **новая концепция смысла** заключается в следующем.

В акте коммуникации двух интеллектов взаимодействуют разные ментальные структуры отображения мира. Успех общения зависит от двух факторов: «взаимопонимания», то есть, от «пересечения» баз знаний индивидуумов, и от «изложения», то есть, от вербальной реализации замыслов коммуникантов. Нас интересует решение поставленной задачи – семантический анализ текстов, иначе говоря – изложение мысли. Поэтому объектом исследования будет тот параметр **текста ЕЯ**, который существенно отображает **тему** общения. Этот параметр назовём **проявленным смыслом**. Проявленный смысл доступен изучению экспертами, в том числе – компьютерному анализу. Такое изучение, или анализ, возможны только на языке репрезентации общих знаний. Машинные знания представлены онтологией ИО\*3, поэтому уместно ввести понятие **онтологический смысл**. В принципе, проявленный смысл, его экспертное толкование и онтологический смысл – не совпадают, но расхождение должно сводиться к минимуму.

Что же представляет собой онтология с позиций представления смысла? Общая аксиома, что «знания несут в себе смысл» обоснована коренным опытом общественного развития. Остаётся уточнить и формализовать связь между ними.

Подчеркнём, что концептами онтологии служат слова естественного языка. Для носителя языка названия предметов и действий имеют **изначальный смысл** даже вне контекста ситуации. Это значит, что при слове «кувшин» у нас сразу возникает понимание, «что это такое», хотя без продолжения оно лишено более глубокого – **контекстного смысла**. Образ кувшина давно оформлен в нашей памяти в виде имени (лексемы), образа и атрибутов предмета; он составляет «матрицу» личного опыта и в то же время является элементом базы знаний человеческого общества. Понятие «кувшин» содержит узуальный (традиционный) смысл.

Тривиальность данных рассуждений приводит к нетривиальному выводу: концепты онтологии должны выбираться из множества осмысленных элементов знания. Понятие «осмысленный элемент знания» распространяется и на такие понятия, которые неочевидны, но каждое из них имеет смысловое толкование либо в словаре ЕЯ, либо в библиотечных источниках и учебниках. В принципе, смысловая лексема должна быть понятна пользователю без ментального напряжения.

Таким образом, в нашей, онтологической концепции **за каждым онтологическим термином скрывается изначальный смысл**. Этот смысл поддерживается базой знаний человечества и в ней интерпретируется.

---

## Формализация

---

1. Введём понятие **элемент смысла**, или **конструкт онтологии**. Под этим будем понимать пару разных лексем: **A** и **B**, которые между собой связаны. **Формат** записи конструкта онтологии: это граф с двумя вершинами: **A—B**. Каким образом связаны узлы – пока не уточняется. Конструкт можно понимать как двуместный предикат: «**A связано с B**».

2. К элементу смысла может быть присоединён другой конструкт: **A—C**, который имеет с первым одинаковый узел. Два одинаковых узла сливаются в один. Получается лексическая цепочка: (**A—B—C**), либо лексический граф: (**A—B, A—C**). Процесс многократного присоединения порождает **онтологическую сеть (граф)** и ограничен множеством понятий ЕЯ. Граф онтологии является направленным, однако допускаются локальные циклические замыкания: **B—C**.

3. **Семантической траекторией** называется подграф онтологического графа, который начинается в некотором узле и заканчивается в вершине онтологии. Семантическая траектория осуществляет дискретный переход от начального слова к конечному.

Используя формальные предпосылки и определение «элемента смысла» как конструкта онтологии: **A—B**, дадим следующее определение понятия «проявленный смысл», который отображён в текстовом документе.

***Проявленным смыслом некоторого связанного текста является совокупность семантических траекторий, каждая из которых начинается в вершине, активизированной ключевым словом текста («связанном концепте») и является подграфом онтологического графа.***

---

## Задача семантического анализа текста.

---

Задана концептуальная онтология и система интерпретации концептов и связей.

Задан текст ЕЯ и совокупность ключевых слов данного текста, которые выделены лингвистическим процессором в качестве значимых слов дискурса. Ключевые слова активизируют связанные концепты.

**Для каждого связанного концепта онтологии необходимо найти на ориентированном онтологическом графе все семантические траектории, которые начинаются в связанном**

---

*концепте, и среди них выделить одну траекторию ограниченной длины, проходящую через максимальное число концептов онтологии.*

*Для заданного текста требуется по совокупности ключевых слов построить совокупность выделенных семантических траекторий онтологического графа.*

Задача приведена в общей формулировке. Уместно привести некоторые пояснения.

1. Каждая семантическая траектория является проекцией текста на онтологические знания и лексически выраженные смыслы. Она непосредственно указывает на то, какие слова выбрал автор текста для выражения своей мысли и как он их связал между собой.

2. В процессе анализа текста допускается пополнение онтологии новыми концептами и связями. При этом стрелкам графа могут быть присвоены значения, взятые непосредственно из текста. Тогда конструкт **А — В** превратится в предикат: «**А –действие– В**». Например: «Субъект – убил – Брат». Благодаря данной процедуре формально-онтологический смысл текста может быть существенно углублен и лучше понят.

---

### Аналогии и результаты

---

Предложенный формально-онтологический подход к обнаружению проявленного в тексте смысла предполагает ряд операций, которые могут быть сопоставлены с *процессами познания* сущностей. Действительно, как следует из нейрофизиологических и нейролингвистических представлений [7, 8], человек выделяет сущность *ситуации* и формирует её *смысл*. Сущность ситуации ассоциируется с системой возбуждённых ментальных моделей – внутренней базой знаний, и если готовая смысловая структура (матрица) имеется, то она «узнаётся» и не меняется. В противном случае в памяти человека, в системе моделей среды, возникают новые связи, новые понятия и кластеры понятий. Их можно понимать как новые знания.

Если проводить функциональные параллели, то онтология служит внешней базой знаний. Алгоритм семантического анализа находит в тексте ключевые слова, которые проявляют «сущность» ситуации. Если слово «узнаётся» онтологией, то активизируется адекватный ему концепт, который, в свою очередь, «возбуждает» соседние концепты. Возникает «цепная реакция» соединения «изначальных», контекстно-свободных смыслов (заданных своими именами) в связную систему – семантическую траекторию. Семантическая траектория организует «понимание» как смысловую интерпретацию текста в контексте общих онтологических знаний.

Если облечь выделенную графическую структуру в грамматически правильное утверждение, то получим **онтологический смысл ключевого слова**. Совокупность траекторий, которая получена в результате семантического анализа полного текста, является **интегральным онтологическим смыслом текста**.

Связность семантического толкования заданного текста (при условии, что лингвистический процессор правильно выбрал ключевые слова) гарантируется связями внутри онтологии.

Ключевые слова текста, которые в онтологии не проявлены, но имеют предикативные связи с другими словами, проявленными в онтологии, могут быть закреплены в структуре новыми концептами и связями.

---

### Выводы

---

Новый подход к формальному представлению смысла базируется на прямой связи смысла с общими знаниями человека о мире. Знания представлены концептуальной онтологией, а элементарные смыслы представлены конструктами сети – двуместными предикатами. За каждым концептом закреплён его

именованный (изначальный) смысл, который без ментального напряжения ясен человеку, и в то же время в декларативной форме он раскрывается в опубликованных источниках – интерпретаторах знаний. Процедура семантического анализа ЕЯ текстов выглядит как «прорастание» в онтологической сети связей, активизированных ключевыми словами текста. Решением задачи контекстного анализа значимого слова является подграф онтологического графа, названный семантической траекторией, а результатом семантического анализа всего текста служит совокупность траекторий. Семантическая траектория реконструирует проявленный в тексте смысл в формате дискретно-концептуального представления, которое затем (при необходимости) может быть преобразовано в формат естественного языка.

---

### Благодарности

---

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теории и Приложений FOI ITHEA и Консорциума FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

### Литература

---

1. Палагін О.В., Петренко М.Г. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи // Математичні машини і системи. – 2007. – №1. – с. 42–50.
2. В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов. Advanced Research in Artificial Intelligence. Supplement to International Journal "Information Technologies and Knowledge" Volume 2 / 2008. – ITHEA, Sofia, 2008. – p. 153.
3. Мельчук И.А. Опыт теории лингвистических моделей «Смысл–Текст». – М.: Школа «языки рус. культуры», 1999.– 346 с.
4. Хокинс Дж., Блейкли С. Об интеллекте: пер. с англ. – М.: ООО «И.Д. Вильямс», 2007. – 240 с.
5. Владимир Лефевр. Алгебра совести: пер. с англ. – М.: Когито-Центр, 2003. – 173 с.
6. Демьянков В.З. Основы теории интерпретации и её приложения в вычислительной лингвистике. – М.: Изд-во Моск. ун-та, 1985. – 76 с.
7. Мельников Г.П. Системология и языковые аспекты кибернетики. – М.: «Сов. радио», 1978. – 368 с.
8. Кисленко Ю.И. От мысли к знанию (нейрофизиологические основания): Монография. – Киев, Изд-во «Український літопис», 2008. – 101 с.

---

### Информация об авторах

---

**Гладун Виктор Поликарпович** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [aduis@rambler.ru](mailto:aduis@rambler.ru)

**Святогор Леонид Александрович** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [aduis@rambler.ru](mailto:aduis@rambler.ru)





---

---

## Neural Nets

---

---

### SELF-LEARNING FUZZY SPIKING NEURAL NETWORK AS A NONLINEAR PULSE-POSITION THRESHOLD DETECTION DYNAMIC SYSTEM BASED ON SECOND-ORDER CRITICALLY DAMPED RESPONSE UNITS

**Yevgeniy Bodyanskiy, Artem Dolotov, Iryna Pliss**

**Abstract:** Architecture and learning algorithm of self-learning spiking neural network in fuzzy clustering task are outlined. Fuzzy receptive neurons for pulse-position transformation of input data are considered. It is proposed to treat a spiking neural network in terms of classical automatic control theory apparatus based on the Laplace transform. It is shown that synapse functioning can be easily modeled by a second order damped response unit. Spiking neuron soma is presented as a threshold detection unit. Thus, the proposed fuzzy spiking neural network is an analog-digital nonlinear pulse-position dynamic system. It is demonstrated how fuzzy probabilistic and possibilistic clustering approaches can be implemented on the base of the presented spiking neural network.

**Keywords:** computational intelligence, hybrid intelligent system, spiking neural network, fuzzy receptive neuron, fuzzy clustering, automatic control theory, analog-digital system, second order damped response system.

**ACM Classification Keywords:** 1.2.6 [Artificial Intelligence]: Learning – Connectionism and neural nets; 1.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – Control theory; 1.5.1 [Pattern Recognition]: Models – Fuzzy set, Neural nets; 1.5.3 [Pattern Recognition]: Clustering – Algorithms.

**Conference:** The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009

---

#### Introduction

---

Among a variety of computational intelligence means for data processing in the absence of a priori information [Haykin, 1999; Sato-Ilic, 2006], self-learning spiking neural networks (SLSNNs) are attracting growing attention both as biologically more realistic models than neural networks of the previous generations [Hopfield, 1995; Gerstner, 2002] and as considerably fast and computationally powerful processing systems [Natschlaeger, 1998; Maass, 1997]. For the last decade, SLSNNs have been successfully used in complex data processing problems solving, particularly in satellite image processing [Bohte, 2002]. Moreover, hybrid intelligent systems combining SLSNNs and fuzzy methodology approaches, known as self-learning fuzzy spiking neural networks (SLFSNNs), revealed a new area where spiking neural networks can be successfully applied, namely fuzzy clustering tasks [Bodyanskiy, 2008a-d].

Although spiking neural networks are becoming a popular computational intelligence tool for various technical problems solving, their architecture and functioning are treated in terms of neurophysiology rather than in terms of any technical sciences apparatus.

In this paper, a technically plausible description of a spiking neural network is introduced. It is proposed to define a spiking neural network in terms of well-known and widely used apparatus of classical automatic control theory based on the Laplace transform. It is shown that a spiking neural network is a pulse-position threshold detection system based on second-order damped response units. Such kind of description allows of, on the one hand, using it as an analog-digital system in technical problems solving. On the other hand, spiking neural network architecture and functioning formalizing simplifies the further spiking neural networks theoretical research.

### Self-Learning Fuzzy Spiking Neural Network Architecture

A self-learning fuzzy spiking neural network is depicted on Figure 1. As illustrated, it is a heterogeneous three-layered feed-forward neural network with lateral connections in the second hidden layer.

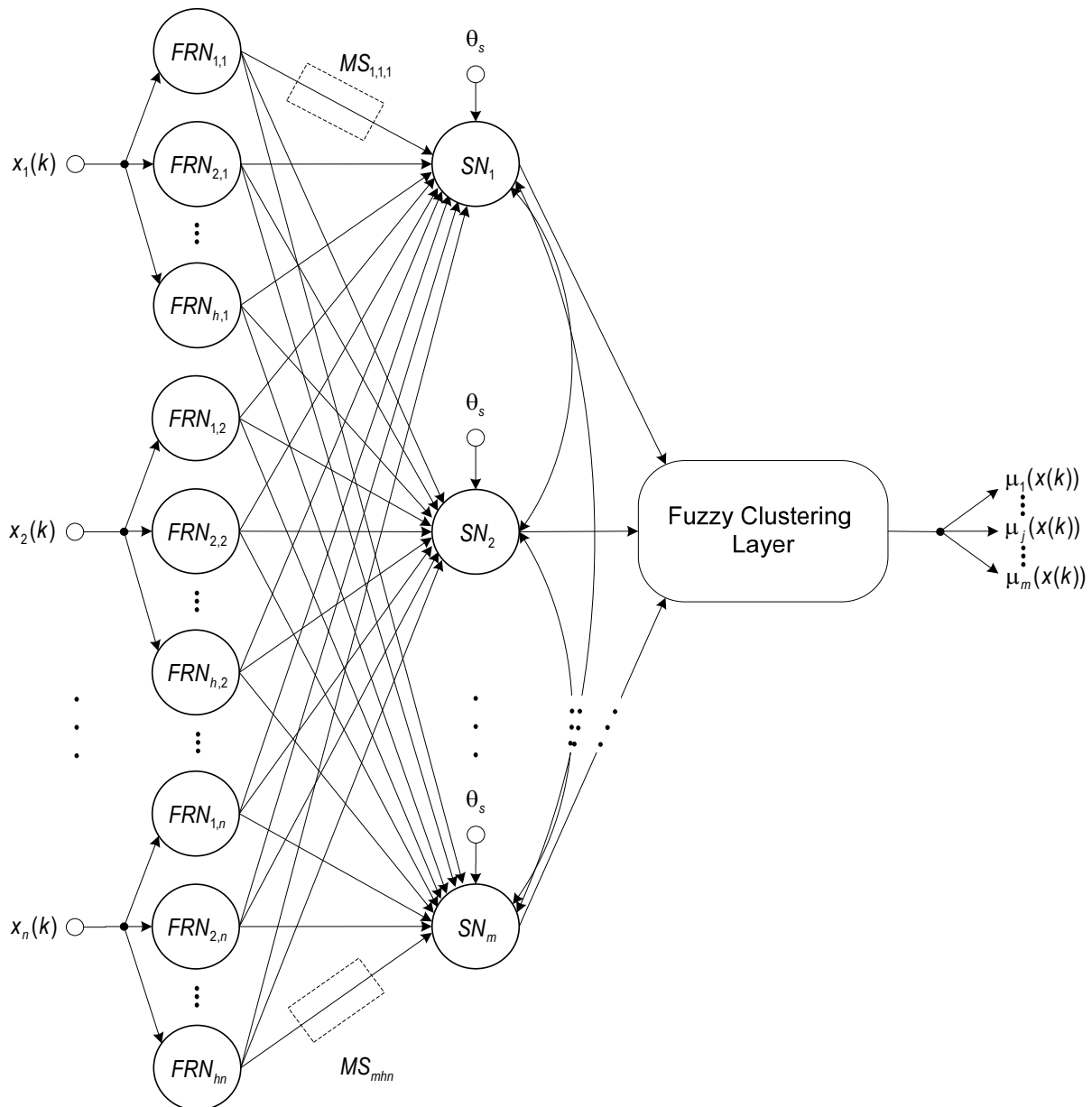


Figure 1. Self-learning fuzzy spiking neural network architecture

The first hidden layer performs pulse-position transformation of  $n \times 1$ -dimensional input patterns  $x(k)$  (here,  $k = 0, 1, \dots, N$  is a pattern number) to the input vector of spikes  $\delta(t - t^{[0]}(x(k)))$  where each spike is defined by its

firing time ( $\delta(\bullet)$  is the Dirac delta function). The transformation is performed by population coding that implies that an input  $x_i(k)$ ,  $i = 1, 2, \dots, n$ , is processed at the same time by a pool of  $h$  fuzzy receptive neurons  $FRN_{li}$ ,  $l = 1, 2, \dots, h$ .

Clusters detection takes place in the second hidden layer that consists of  $m$  spiking neurons  $SN_j$ ,  $j = 1, 2, \dots, m$  ( $m$  is a number of clusters to be detected). They are connected with neurons of the previous layer by multiple synapses  $MS_{jli}$ . After learning phase, a spiking neuron  $SN_j$  emits outgoing spike  $\delta(t - t_j^{[1]}(x(k)))$  for each input pattern  $x(k)$ , and the neuron firing time defines the distance of the input pattern to the neuron's center.

The third layer processes distances of the input patterns to spiking neurons' centers, performs fuzzy partitioning, and produces the membership levels  $\mu_j(x(k))$ ,  $j = 1, 2, \dots, m$ .

It is worth to note that two first hidden layers form conventional architecture of SLSNN [Bohte, 2002]. In case of such network using, the cluster that an input pattern belongs to is determined by the earliest fired spiking neuron.

---

### Fuzzy Receptive Neurons

---

Architecture of fuzzy receptive neurons of the first hidden layer [Bodyanskiy, 2008c] is identical to the one of receptive neurons that were proposed to perform population coding in SLSNNs [Bohte, 2002]. The difference between them is an interpretation of their functioning and the method of activation functions setting.

As a rule, a receptive neuron activation function is bell-shaped (Gaussian usually), and activation functions of the neurons within a pool are shifted, overlapped, and of different widths. In a general case, firing time of a spike emitted by receptive neuron lies in a certain interval  $[0, t_{max}^{[0]}]$  referred to as a coding interval and is defined by the expression

$$t_{li}^{[0]}(x_i(k)) = \lfloor t_{max}^{[0]} \left( 1 - \psi \left( \left| x_i(k) - c_{li}^{[0]} \right|, \sigma_i \right) \right) \rfloor, \quad (1)$$

where  $\lfloor \bullet \rfloor$  is the floor function,  $\psi(\bullet, \bullet)$ ,  $c_{li}^{[0]}$ , and  $\sigma_i$  are the receptive neuron's activation function, center, and width respectively.

One can readily see that the layer of receptive neurons pools is identical to a fuzzification layer of neuro-fuzzy systems like Takagi-Sugeno-Kang networks, ANFIS, etc [Jang, 1997]. Considering activation function  $\psi_{li}(x_i(k))$  as a membership function, the receptive neurons layer can be treated as the one that transforms input data set to a fuzzy set that is defined by values of activation-membership function  $\psi_{li}(x_i(k))$  and is expressed over time domain in form of firing times  $t_{li}^{[0]}(x_i(k))$ . In fact, each pool of receptive neurons performs zero order Takagi-Sugeno fuzzy inference [Jang, 1997]

$$\text{IF } x_i(k) \text{ IS } X_{li} \text{ THEN OUTPUT IS } t_{li}^{[0]}, \quad (2)$$

where  $X_{li}$  is the fuzzy set with membership function  $\psi_{li}(x_i(k))$ . Thus, one can interpret a receptive neurons pool as a certain linguistic variable and each receptive neuron (more precisely, fuzzy receptive neuron) within the pool – as a linguistic term with membership function  $\psi_{li}(x_i(k))$ . This way, having any a priori knowledge of data structure, it is possible to adjust activation functions of the first layer neurons to fit them and thus, to get better clustering results.

---

### Spiking Neuron as a Nonlinear Dynamic System

---

Spiking neuron as a nonlinear dynamic system is depicted on Figure 2. As illustrated, multiple synapses of spiking neuron  $SN_j$  transform the incoming pulse-position signal to a continuous-time form, and its soma transforms the incoming continuous-time signal back to pulse-position form.

In a scope of automatic control theory [Goodwin, 2001; Phillips, 2000; Dorf, 2001], multiple synapse  $MS_{jji}$  is a dynamic system that consists of different time delays, second-order damped response units, and adjustable gains that are connected in parallel. Each group of time delay, second-order damped response unit, and gain form a subsynapse of multiple synapse. As a response to incoming spike, the subsynapse produces delayed weighted postsynaptic potential  $u_{jji}^p(t)$ , and the multiple synapse produces total postsynaptic potential  $u_{jji}(t)$  that arrives to spiking neuron soma.

Transfer function of a second-order damped response unit with unit gain factor is

$$G(s) = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)} = \frac{1}{\tau_4^2 s^2 + \tau_3 s + 1}, \quad (3)$$

where  $s$  is the Laplace operator,  $\tau_{1,2} = \frac{\tau_3}{2} \pm \sqrt{\frac{\tau_3^2}{4} - \tau_4^2}$ ,  $\tau_1 \geq \tau_2$ ,  $\tau_3 \geq 2\tau_4$ , and its impulse response is

$$\tilde{\varepsilon}(t) = \frac{1}{\tau_1 - \tau_2} \left( e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}} \right). \quad (4)$$

Putting  $\tau_1 = \tau_2$  (that corresponds to a second-order critically damped response system) and applying l'Hôpital's rule, one can obtain

$$\tilde{\varepsilon}(t) = \frac{t}{\tau_1^2} e^{-\frac{t}{\tau_1}} \quad (5)$$

Comparing a spike-response function [Gerstner, 2002]

$$\varepsilon(t) = \frac{t}{\tau} e^{1-\frac{t}{\tau}}, \quad (6)$$

where  $\tau$  is the membrane potential decay time constant, with (5) leads us to the following expression:

$$\varepsilon(t) = e\tau\tilde{\varepsilon}(t). \quad (7)$$

Thus, transfer function of the second-order critically damped response unit whose impulse response corresponds to a spike-response function is

$$G(s) = \frac{e\tau}{(\tau s + 1)^2}. \quad (8)$$

Taking into account (8), transfer function of the  $p$ -th subsynapse of  $MS_{jji}$  takes form

$$U_{jji}^p(s) = \frac{\tau w_{jji}^p e^{1-d^p s}}{(\tau s + 1)^2}, \quad (9)$$

where  $w_{jji}^p$  and  $d^p$  are synaptic weight and time delay of the subsynapse.

The Laplace transform of a spike  $\delta(t - t_{ji}^{[0]}(x_i(k)))$  is

$$L\{\delta(t - t_{ji}^{[0]}(x_i(k)))\} = e^{-t_{ji}^{[0]}(x_i(k))s}, \quad (10)$$

so taking into account transfer function of multiple synapse  $MS_{jji}$

$$U_{jji}(s) = \sum_{p=1}^q U_{jji}^p(s) = \sum_{p=1}^q \frac{\tau w_{jji}^p e^{1-d^p s}}{(\tau s + 1)^2}, \quad (11)$$

where  $q$  is a number of subsynapses within a multiple synapse, the Laplace transform of the multiple synapse output can be expressed in the following form:

$$u_{jji}(s) = e^{-t_{ji}^{[0]}(x_i(k))s} U_{jji}(s) = e^{-t_{ji}^{[0]}(x_i(k))s} \sum_{p=1}^q \frac{\tau w_{jji}^p e^{1-d^p s}}{(\tau s + 1)^2} = \sum_{p=1}^q \frac{\tau w_{jji}^p e^{1-(t_{ji}^{[0]}(x_i(k))+d^p)s}}{(\tau s + 1)^2}. \quad (12)$$

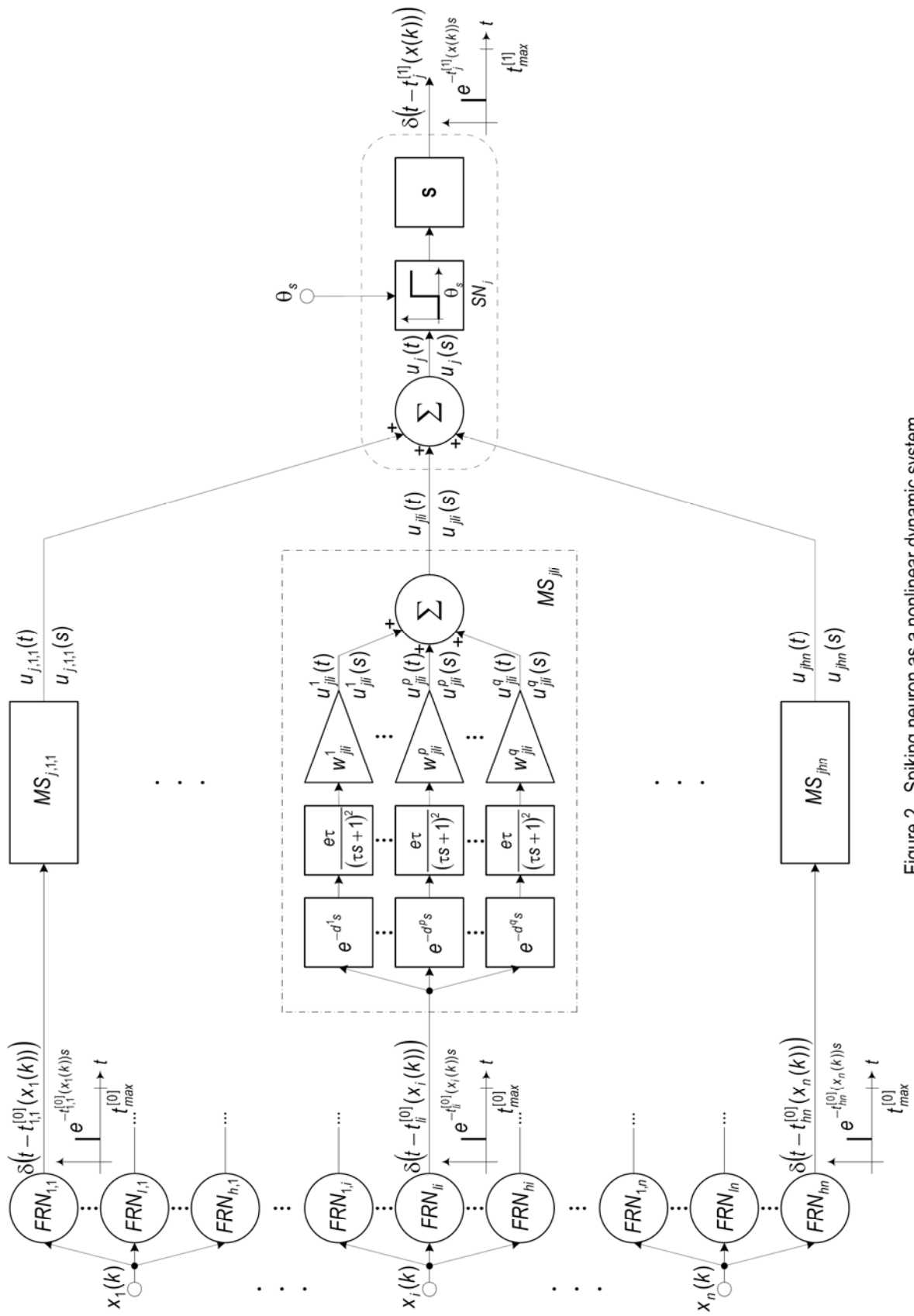


Figure 2. Spiking neuron as a nonlinear dynamic system

Here it is worth to note that since it is impossible to use  $\delta$ -function in practice [Phillips, 2000], it is convenient to model it with impulse of a triangular form as shown on Figure 3. Such impulse is similar to a biological spike and satisfies the condition

$$\lim_{\Delta \rightarrow 0} \rho(t, \Delta) = \delta(t). \quad (13)$$

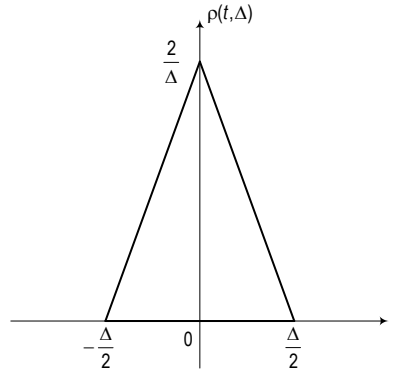


Figure 3. Triangular impulse

The Laplace transform of spiking neuron  $SN_j$  membrane potential can be expressed as follows:

$$u_j(s) = \sum_{i=1}^n \sum_{l=1}^h \sum_{p=1}^q \frac{\tau w_{jli}^p e^{1-(t_i^{[0]}(x_i(k))+d^p)s}}{(\tau s + 1)^2}. \quad (14)$$

Spiking neuron soma firing behavior is modeled by an element relay with dead zone  $\theta_s$  that is defined by nonlinear function

$$f(u) = \frac{\text{sign}(u - \theta_s) + 1}{2}, \quad (15)$$

and a derivative unit with transfer function

$$G_D(s) = s \quad (16)$$

being connected in series.

At the instance when soma membrane potential  $u_j(t)$  reaches the firing threshold  $\theta_s$ , the element relay triggers and emits the Heaviside step functions on its output. Differentiating the latter gives an outgoing spike  $\delta(t - t_j^{[1]}(x(k)))$ . Thus, spiking neuron soma functions as a threshold detection unit.

During learning phase, on each learning epoch, the temporal Hebbian rule updates weights of the spiking neuron-winner in the following way [Natschlaeger, 1998; Bohte, 2002]: the weights of those subsynapses which contributed to the neuron's firing are strengthened, whereas weights of subsynapses which did not contribute are weakened. Thus, weights are adjusted to move the center of the neuron-winner closer to input pattern. Lateral inhibitory connections in the second hidden layer are used only during the learning to implement 'winner-takes-all' mechanism. After learning phase is complete, the lateral connections are disabled.

---

### Output Fuzzy Clustering Layer

---

The output layer, namely output fuzzy clustering layer, takes firing times of spikes  $\delta(t - t_j^{[1]}(x(k)))$  arriving from the second layer, and either performs fuzzy partitioning of the input patterns  $x(k)$  using probabilistic approach [Bodyanskiy, 2008a, b]



$$\mu_j(x(k)) = \frac{(t_j^{[1]}(x(k)))^2}{\sum_{i=1}^m (t_i^{[1]}(x(k)))^2}, \quad (17)$$

where  $\zeta$  is the fuzzifier that determines boundary between clusters and controls the amount of fuzziness in the final partition, or possibilistic approach [Bodyanskiy, 2008d]

$$\mu_j(x(k)) = \left( 1 + \left( \frac{(t_j^{[1]}(x(k)))^2}{\lambda_j} \right)^{\frac{1}{\zeta-1}} \right)^{-1}, \quad (18)$$

$$\lambda_j = \frac{\sum_{k=1}^N \mu_j^\zeta (t_j^{[1]}(x(k)))^2}{\sum_{k=1}^N \mu_j^\zeta (x(k))}, \quad (19)$$

It is readily seen that the output layer evaluates fuzzy membership similarly to well-known fuzzy c-means or possibilistic c-means algorithms [Bezdek, 2005] – depending on the used approach.

Output fuzzy clustering layer is disabled during learning phase and is used on classification phase only.

---

## Conclusion

Spiking neural networks are more realistic models of real neuronal systems than artificial neural networks of the previous generations. Nevertheless, they can be described in a strict technically plausible way. Treating a spiking neural network in a scope of automatic control theory, it is easily seen that spiking neuron synapse is nothing other than a second-order damped response system, and the soma is a threshold detection system. Spiking neural network implemented on their basis is an analog-digital nonlinear dynamic system that conveys and processes information both in pulse-position and continuous-time forms. Such precise formal description of spiking neural network architecture and functioning presents a significant step toward evolving of artificial neural networks theory as a part of computational intelligence paradigm.

---

## Acknowledgement

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

## Bibliography

- [Bezdek, 2005] J.C.Bezdek, J.Keller, R.Krishnapuram, N.R.Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Springer, New York, 2005.
- [Bodyanskiy, 2008a] Ye.Bodyanskiy, A.Dolotov. Image processing using self-learning fuzzy spiking neural network in the presence of overlapping classes. Proc. 11th Int. Biennial Baltic Electronics Conf. "BEC 2008" (October 6-8, 2008, Tallinn, Estonia), Tallinn University of Technology, Tallinn, 2008, P. 213-216.
- [Bodyanskiy, 2008b] Ye.Bodyanskiy, A.Dolotov. A Self-learning spiking neural network for fuzzy clustering task. Sci. Proc. of Riga Technical University: Information Technology and Management Science, 36, 2008, P. 27-33.

- [Bodyanskiy, 2008c] Ye.Bodyanskiy, A.Dolotov, I.Pliss. Fuzzy receptive neurons using in self-learning spiking neural network. Proc. Int. Sci. and Tech. Conf. "Automation: Problems, Ideas, Solutions" (September 8-12, 2008, Sevastopol, Ukraine), Sevastopol National Technical University, Sevastopol, 2008, P. 12-14 (in Russian).
- [Bodyanskiy, 2008d] Ye.Bodyanskiy, A.Dolotov, I.Pliss, Ye.Viktorov. Fuzzy possibilistic clustering using self-learning spiking neural network. Wissenschaftliche Berichte der Hochschule Zittau/Goerlitz, Heft 100, Nr. 2360-2395, 2008, S. 53-60.
- [Bohte, 2002] S.M.Bohte, J.S.Kok, H.La Poutre. Unsupervised clustering with spiking neurons by sparse temporal coding and multi-layer RBF networks. IEEE Trans. on Neural Networks, 13, 2002, P. 426-435.
- [Dorf, 2001] P.C.Dorf, R.H.Bishop. Modern Control Systems. Prentice Hall, Upper Saddle River, 2001.
- [Gerstner, 2002] W.Gerstner, W.M.Kistler. Spiking Neuron Models. The Cambridge University Press, Cambridge, 2002.
- [Goodwin, 2001] G.C. Goodwin, S.F.Graebe, M.E.Salgado. Control Systems Design. Prentice Hall, Upper Saddle River, 2001.
- [Haykin, 1999] S.Haykin. Neural Networks. A Comprehensive Foundation. Prentice Hall, Upper Saddle River, 1999.
- [Hopfield, 1995] J.J.Hopfield. Pattern recognition computation using action potential timing for stimulus representation. Nature, 376, 1995, P. 33-36.
- [Jang, 1997] J.-S.R.Jang, C.-T.Sun, E.Mizutani. Neuro-Fuzzy and Soft Computing. Prentice Hall, Upper Saddle River, 1997.
- [Maass, 1997] W.Maass. Networks of spiking neurons: the third generation of neural network models. Neural Networks, 10, 1997, P. 1659-1671.
- [Natschlaeger, 1998] T.Natschlaeger, B.Ruf. Spatial and temporal pattern analysis via spiking neurons. Network: Computations in Neural Systems, 9, 1998, P. 319-332.
- [Phillips, 2000] C.L.Phillips, R.D.Harbor. Feedback Control Systems. Prentice Hall, Upper Saddle River, 2000.
- [Sato-Ilic, 2006] M.Sato-Ilic, L.C.Jain. Innovations in Fuzzy Clustering. Springer, Berlin – Heidelberg – New York, 2006.
- 

### **Authors' Information**

---

**Yevgeniy Bodyanskiy** – Professor, Dr.-Ing. habil., Scientific Head of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine;  
e-mail: [bodya@kture.kharkov.ua](mailto:bodya@kture.kharkov.ua).

**Artem Dolotov** – Ph. D. student, Junior Researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine;  
e-mail: [dolotov.artem@gmail.com](mailto:dolotov.artem@gmail.com).

**Iryna Pliss** – Ph. D., Leading Researcher of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine;  
e-mail: [pliss@kture.kharkov.ua](mailto:pliss@kture.kharkov.ua).

---

## PERFORMANCE COMPARISON OF MATLAB AND NEURO SOLUTION SOFTWARE ON ESTIMATION OF FUEL ECONOMY BY USING ARTIFICIAL NEURAL NETWORK

**Hakan Serhad Soyhan, Mehmet Emre Kilic, Burak Gokalp, Imdat Taymaz**

**Abstract:** *In the world, scientific studies increase day by day and computer programs facilitate the human's life. Scientists examine the human's brain's neural structure and they try to be model in the computer and they give the name of artificial neural network. For this reason, they think to develop more complex problem's solution. The purpose of this study is to estimate fuel economy of an automobile engine by using artificial neural network (ANN) algorithm. Engine characteristics were simulated by using "Neuro Solution" software. The same data is used in MATLAB to compare the performance of MATLAB is such a problem and show its validity. The cylinder, displacement, power, weight, acceleration and vehicle production year are used as input data and miles per gallon (MPG) are used as target data. An Artificial Neural Network model was developed and 70% of data were used as training data, 15% of data were used as testing data and 15% of data is used as validation data. In creating our model, proper neuron number is carefully selected to increase the speed of the network. Since the problem has a nonlinear structure, multi layer are used in our model.*

**Keywords:** *Artificial Neural Network, Fuel Economy*

**Conference:** *The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009*

---

### 1. Introduction

Since the advent of the diesel engine over a century ago, a huge number of research studies have been conducted to improve the engine performance, decrease the engine fuel consumption and reduce the unwanted exhaust emissions. The oil crisis of the 1970s accelerated the interest of using alternative fuels for diesel engines although diesel fuel has remained as the main fuel [1].

ANN technique has been used for modeling the performance of various thermal systems [2–5]. Recently, ANNs have been applied to estimate various performance parameters of internal combustion engines. This approach was used to predict the performance and exhaust emissions of diesel engines [6,7]. Several studies may be found in different areas of science making use of artificial neural-network [8]. The ANN approach has been applied to predict the performance of various thermal systems. [2]

In this study of a work already in Neuro Solution program using data on fuel economy with the help of Matlab NNTOOL convergence has been made.

---

### 2. Modeling with the ANN

Engine characteristics as the number of cylinders, displacement, power, weight, acceleration and vehicle production year and a miles per gallon (MPG) values as data is used. 70% of the data are the training data values, 15% are the test data and the rest are validation data. Training for data entry in the matlab environment can be transferred after transpose "input" will be saved with the name. Same target values in the "target" will be saved with the name. Test data input and target values "test\_input" and "test\_target" names, validation of data input and target values "validation\_input" and "validation\_target" is saved as. In this study, type of network as feed-forward back prop, education function trainlm as the performance criteria used as the MSE, and this was accepted as constant parameters. Experience and experimentation has been identified as 3, after the number of layers. One will be used between layers of neurons and activation function is the selection of.

The architecture of the ANN for the engine with the input and output parameters is schematically illustrated in Fig.1. The input layer has six neurons corresponding to the three input parameters: cylinder numbers, displacement, engine power, engine weight, acceleration and distance with one-gallon fuel. The number of neurons must be considered in selection. Increase in the number of neurons slows down the neuronal network. This is why we try to capture the optimum number of neurons.

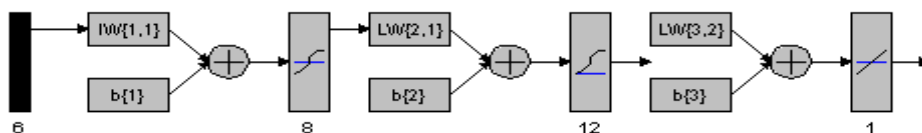


Fig.1. Structure of the ANN for modeling the test

The performance of an ANN model is noticeably affected by the number of hidden layers and the number of nodes in each hidden layer. By trial and error with different ANN configurations, the optimal number of hidden layers and that of neurons in the hidden layer were selected as one and five, respectively. The activation function in the hidden and output layers was chosen as hyperbolic tangent sigmoid function.

In order to develop an ANN model for the engine, experimental data was divided into training data, test sets and validation data. While 70% of the data set was randomly assigned as the training set, the remaining 15% was employed for testing the performance of the ANN predictions. The input vectors with three variables and their corresponding target vectors with six variables in the training set were presented to the network for training it in three separate groups.

The training set of each group consisted of 63 input-output pairs, corresponding to 70% of the data set containing the results of 20 tests in each group. Using the training process was terminated when the maximum number of epochs was exceeded or the performance goal was met. Finally, the input vectors from the test data set were presented to the trained network and the network predictions were compared with the experimental outputs for the performance measurement.

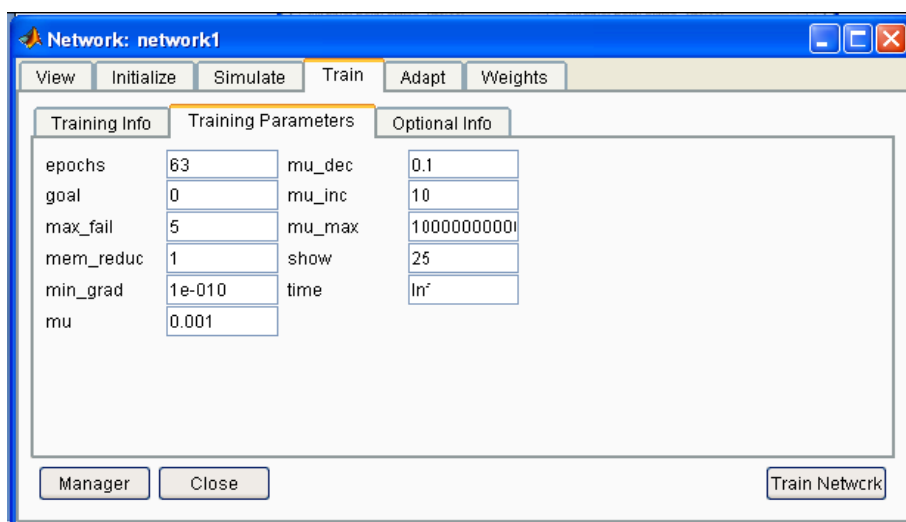


Fig.2. Training parameters are entered

The test data set contained 6 input-output pairs for each group, corresponding to the 15% of the data set containing the results of 63 tests in each group. The computer code for training the ANN and measuring its performance was implemented under MATLAB environment.

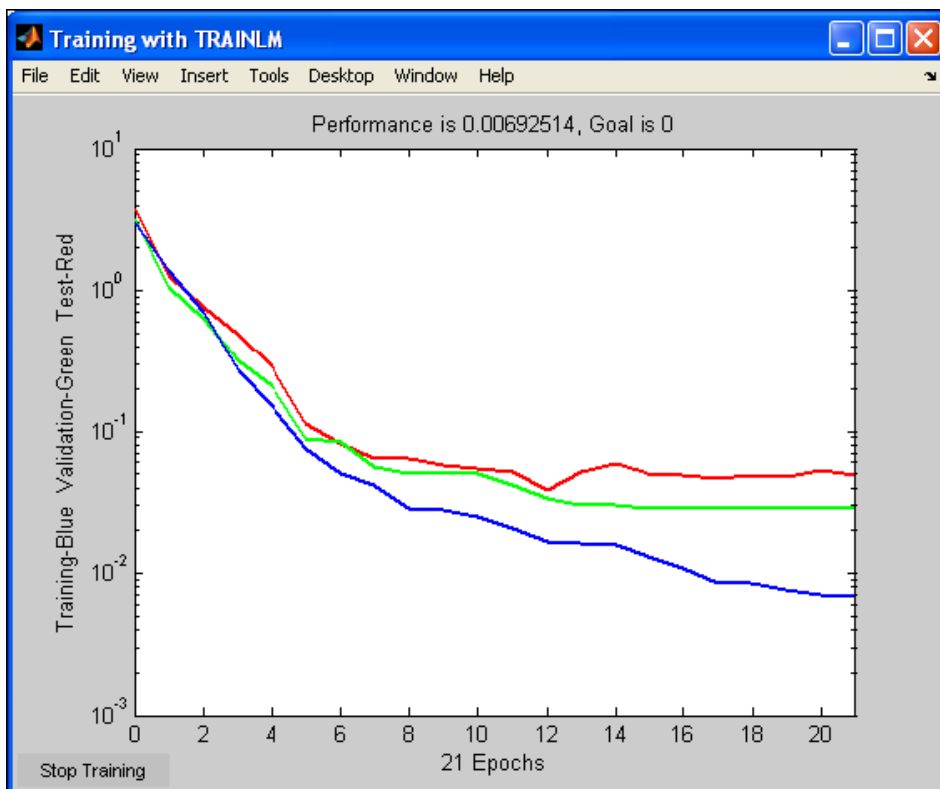


Fig.3. Network error values in Matlab

Under the training curve of Figure 3 test curves that are seen as having less value error. The above network has a problem such as going to memorize. The error in the program for solution of neuron  $2.33601 \times 10^{-4}$  obtained better results have been obtained (Figure 4).

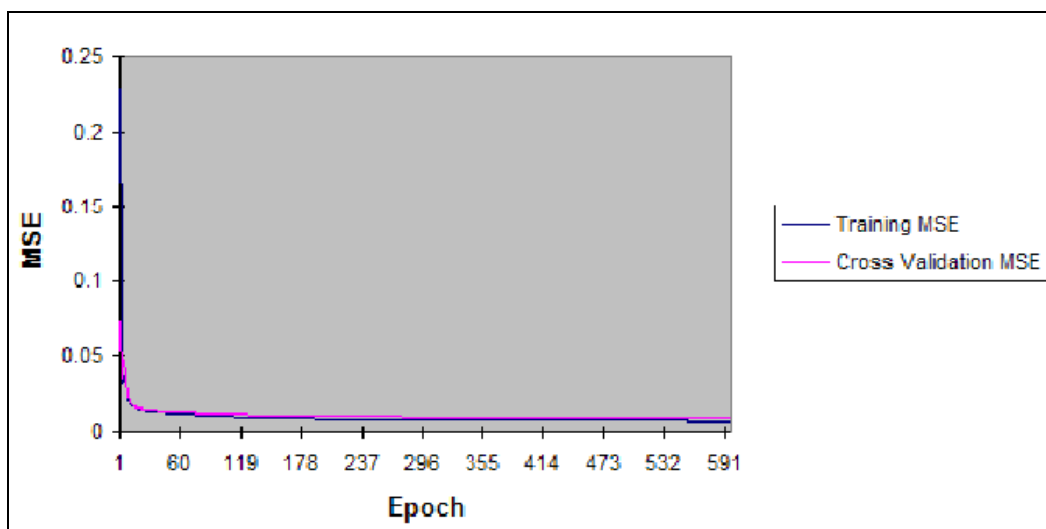


Fig.4. Network error values in Neuro Solution

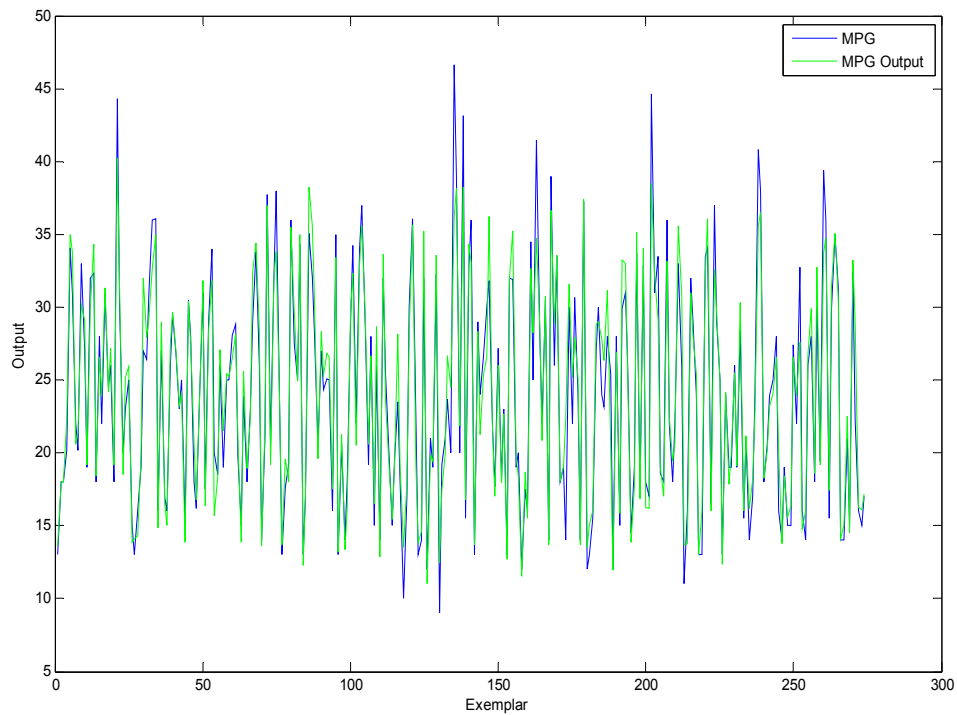


Fig.5. Desired output and actual network output in Matlab

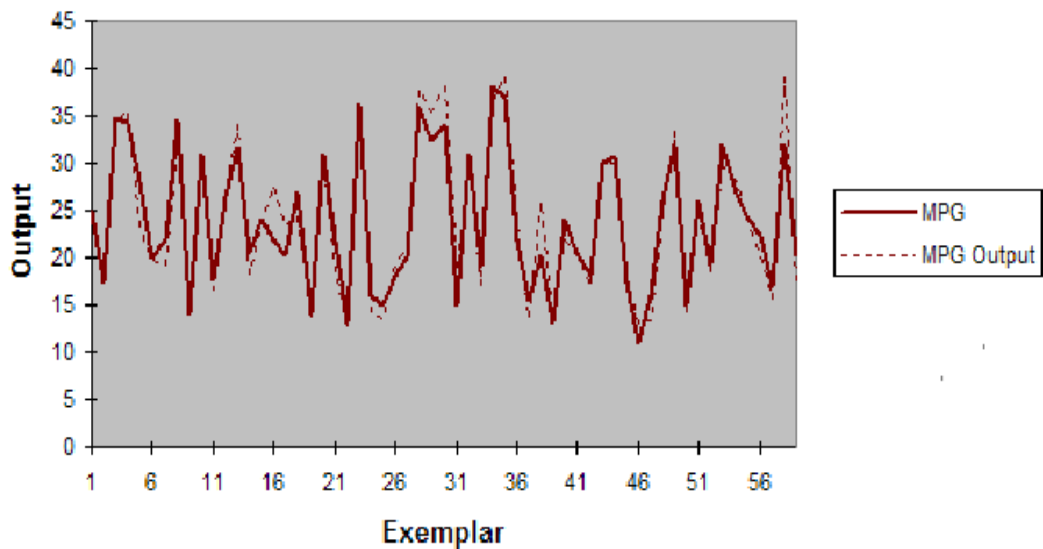


Fig.6. Desired output and actual network output in Neuro Solution

As seen in Figure 5 and figure 6 desired values are reached but some peak points couldn't be reached. It is not possible to reach actual values one to one. Purpose is to reach for the nearest target. This data set excluded from the next peak value is considered, as seen in Figure 7 result in suspension of the target value has been opposed.

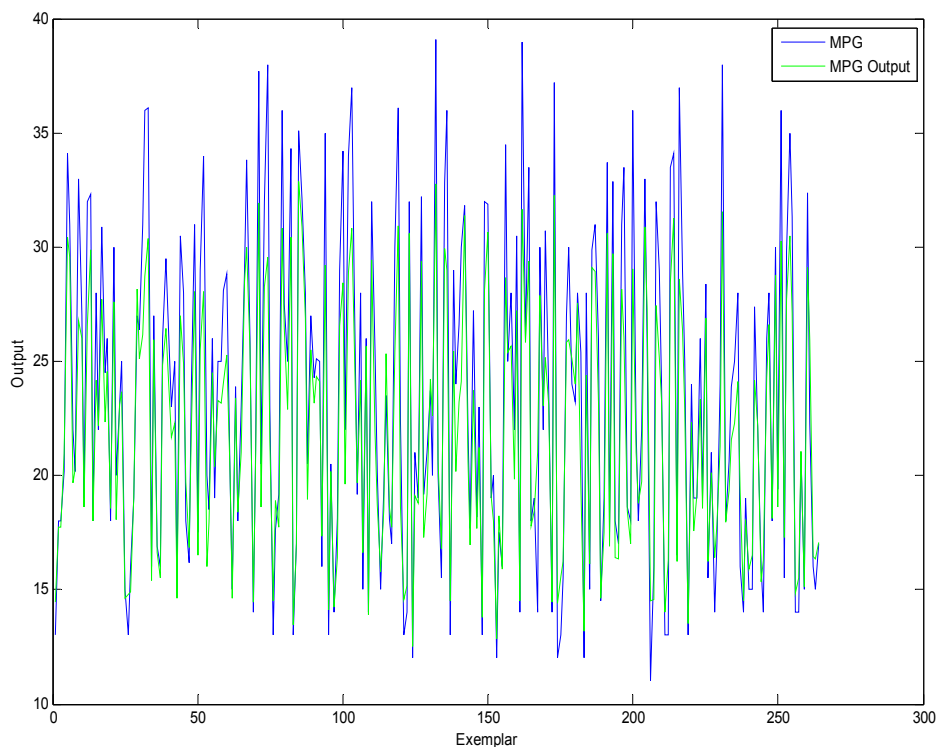


Fig.7. Desired output and actual network output in Matlab after deleted peak values

The ANN is trained, any data input from outside the training set according to the values to generate output. This process is called simulation. During the simulation normalized inputs are simulated in ANN. Be obtained from the value for normalizing the output value will be converted back to normal values are obtained. The following section shows how to make the simulation process:

```
input(:,275)=[8;302;139;3570;12.8;78];
a=premnmx(input);
b=sim(network1,a);
c=postmnmx(b,mintn,maxtn);
result=c(:,275)
result = 19.3208 (in Matlab)
result= 20.2 (in Neuro Solution).
```

Final result have been obtained in Neuro Solution

### 3. Results and Discussion

Comparisons between the ANN-predicted and experimental output parameters of the engine are shown in Figs 1–7. The comparisons in all graphics were made using data only from the test set, which was not introduced to the ANN during the training. In this work we show that ANN can be trained with engine characteristics and target MPG values. Here, we trained ANN to simulate the data set from outside the engine to the engine characteristics by estimating the MPG.

### Acknowledgement

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

**Bibliography**

---

- [1] Canakci M, Van Gerpen JH. Comparison of engine performance and emissions for petroleum diesel fuel, yellow grease biodiesel, and soybean oil biodiesel. *Trans ASAE* 2003;46:937–44.
- [2] Kalogirou SA. Application of artificial neural-networks for energy systems. *Applied Energy* 2000;67:17–35.
- [3] Bechtler H, Browne MW, Bansal PK, Kecman V. Neural networks – a new approach to model vapour-compression heat pumps. *Int J Energy Res* 2001;25:591–99.
- [4] Sozen A, Arcaklioglu E, Ozalp M. A new approach to thermodynamic analysis of ejector-absorbion cycle: artificial neural networks. *Applied Therm Eng* 2003;23:937–52.
- [5] Ertunc HM, Hosoz M. Artificial neural network analysis of a refrigeration system with an evaporative condenser. *Applied Therm Eng* 2006;26:627–35.
- [6] Arcaklioglu E, Celikten I. A diesel engine's performance and exhaust emissions. *Applied Energy* 2005;80:11–22.
- [7] Celik V, Arcaklioglu E. Performance maps of a diesel engine. *Applied Energy* 2005; 81:247–59.
- [8] Arcaklioglu E., Çavuşoğlu A., Erişen A. Thermodynamic Analysis of refrigerant mixtures using artificial neural networks. *Applied Energy* 2004;78:219-30.

---

**Authors Information**

---

**Hakan Serhad Soyhan** – Department of Mechanical Engineering, Sakarya University, 41040 Sakarya, Turkey;  
e-mail: [hsoyhan@hotmail.com](mailto:hsoyhan@hotmail.com)

**Mehmet Emre Kilic** – Department of Mechanical Engineering, Sakarya University, 41040 Sakarya, Turkey

**Burak Gokalp** – Department of Mechanical Engineering, Kocaeli University, 41040 Kocaeli, Turkey

**Imdat Taymaz** – Department of Mechanical Engineering, Sakarya University, 41040 Sakarya, Turkey



---

## THE USAGE OF NEURAL NETWORKS FOR THE MEDICAL DIAGNOSIS

**Kateryna Malyshevskaya**

**Abstract:** *The problem of cancer diagnosis from multi-channel images using the neural networks is investigated. The goal of this work is to classify the different tissue types which are used to determine the cancer risk. The radial basis function networks and backpropagation neural networks are used for classification. The results of experiments are presented.*

**Keywords:** *neural networks, backpropagation, RBF, uterine cervix, cancer, classification.*

**ACM Classification Keywords:** *I.5.1 Pattern Recognition - Neural nets*

**Conference:** *The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009*

---

### Introduction

The possibility of uterine cervix cancer diagnosis is considered in this work. Such problem appeared from the necessity of early diagnostic of disease using the computer system, which would help the doctor to define the tissues with the high risk of the cancer transformed tissue appearance. The system is based on the assumption that the optical characteristics of healthy cells and diseased cells differ and this difference is more significant than variations in such characteristics among cells that belong to different people. At medical university of Arizona (USA) the optical system which in addition to the usual colposcopy testing provides multichannel images of uterine cervix tissue was introduced. Multichannel images of 108 patients were produced. Simultaneously, the same patients were examined by a doctor. The examination consisted of biopsy of certain tissue areas which were sent to a pathologist for an analysis. Biopsy areas were defined on the image and compared with its results. This information was used to develop the risk areas recognition algorithm based on comparison of a multichannel picture, biopsy results, and doctor's diagnosis. Based on obtained information, it is possible to create the system that allows to perform early diagnostics [Schoonmaker J. 2007].

---

### Statement of the problem

According to the medical statistics, the uterine cervix cancer takes the fourth place among women oncological diseases, (after the stomach, skin, and breast cancer). Primary colposcopy examination of the patient defines the necessity of making biopsy and the further consultation of the oncologist [Воробьева Л. И. 2008]. The computer system considered in this work can be used by a doctor for preliminary diagnostics of a cancer by determining the presence of certain tissue types without conducting a biopsy. The European Expert Group developing the European generalized training program, has offered the following classification of UC epithelium transformation: SEA (squamous epithelial abnormalities) - benign changes of the flat epithelium, columnar without the changes, allowing to assume CIN, Squamous cell changes - changes flat epithelium without accurate signs of a tumor, CIN-I - the least risky type, represents only mild dysplasia, or abnormal cell growth, CIN-II - moderate dysplasia confined to the basal 2/3 of the epithelium, CIN-III - severe dysplasia that spans more than 2/3 of the epithelium, and may involve the full thickness (CIS - carcinoma in situ).[ Koss L.G. 1989] The purpose of this work is the development of computer system which can correctly classify different tissue types (SEA, CIN-I, II, III) by multichannel images using neural networks (NN).

## Methods

**Backpropagation neural networks** employ one of the most popular neural network learning algorithms, the Backpropagation (BP) algorithm. It has been used successfully for wide variety of applications, such as speech or voice recognition, image pattern recognition, medical diagnosis, and automatic controls.

The backpropagation algorithm trains a given feed-forward multilayer neural network for a given set of input patterns with known classifications. When each entry of the sample set is presented to the network, the network examines its output response to the sample input pattern. The output response is then compared to the known and desired output and the error value is calculated. Based on the error, the connection weights are adjusted. The backpropagation algorithm is based on Widrow-Hoff delta learning rule in which the weight adjustment is done through mean square error of the output response to the sample input. [Зайченко Ю.П. 2004]

### RBF Neural Networks

Radial Basis Functions are powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a centre. Radial basis functions have been applied in the area of neural networks where they may be used as a replacement for the sigmoidal hidden layer transfer characteristic in Multi-Layer Perceptrons. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability. Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics and known to correspond to a prior belief in small parameter values (and therefore smooth output functions) in a Bayesian framework. [Domagoj Kovacevic, Sven Loncaric 1997]

### Approximators with local basis

RBF NN looks like perceptron with one hidden layer, carrying out nonlinear reflection

$\mathfrak{R}^d \Rightarrow \mathfrak{R}^m \quad y = \sum_i h_i \phi(w_i, x)$ , being a linear combination of basic functions. But unlike perceptrons where these functions depend on projections to a set of hyperplanes  $\sigma(wx)$ , In the RBF NN functions which depend on distances to basic centers (often Gaussian) are used:  $y = \sum_i h_i \phi_i(|w_i - x|), \phi_i(z) = e^{-z^2 / \sigma_i^2}$ .

Both sets of basic functions provide possibility of approximation of any continuous function with any accuracy. The main distinction between them is the method of information coding on a hidden layer. If perceptrons use global variables (sets of infinite hyperplanes) RBF networks depend on the compact spheres surrounding a set of basic centers (fig. 1).

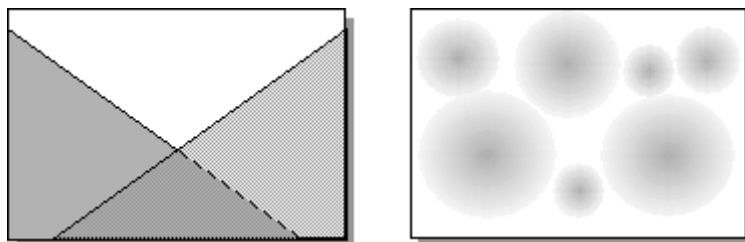


Fig. 1. Global (perceptron) and local (RBF NN) methods of approximation.

In the first case all neurons of hidden layer participate in approximation in a area of any point, in the second case - only the nearest neurons. As consequence of such inefficiency, in the latter case the quantity of the support function necessary for approximating with set accuracy exponentially increases with dimension of space. It is the main disadvantage of networks of radial basis [5].

**Advantages** RBF networks have the advantage of not suffering from local minima in the same way as Multi-Layer Perceptrons. This is because the only parameters that are adjusted in the learning process are the linear

mapping from hidden layer to output layer. Linearity ensures that the error surface is quadratic and therefore has a single easily found minimum. In regression problems this can be found in one matrix operation. In classification problems the fixed non-linearity introduced by the sigmoid output function is most efficiently dealt with using iteratively re-weighted least squares.

### Data preprocessing

Many indicators have high correlation. It can have negative impact on the neural networks work and classification of the tissues on normal/abnormal. For this purpose in order to avoid potential problems in two cases indicators have been changed. In one case, only a subset of indicators with correlation  $<0,95$  was used for neural network training (defined experimentally). In other case, the Principal Component Analysis was used, indicators have been transformed to a set of independent non correlated indicators. [M.C. Jones and R. Sibson 1987]. These methods should improve the functioning of a neural network and increase accuracy of classification.

### Experiments results

There are the diagnoses of two doctors representing presence of each of 6 possible types of tissues (Squamous, Columnar, SEA, CIN1, CIN2, CIN3). The following tables present the results of applying the neural networks. During the experiments, the Cross validation method was used. Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability in the overall assessment of generalizability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

Tables 1 through 6 show the mean square error (MSE) for four types of neural networks (NN), namely Backpropagation Conjugate gradient NN, Backpropagation Quasi Newton NN, Backpropagation Scaled Conjugate gradient NN, and RBF NN each with two methods of data preprocessing (a subset of indicators with correlation  $<0,95$  and the indicators transformed using PCA) for the following tissue types: Squamous, Columnar, SEA, CIN1, CIN2, and CIN3.

**Table 1.** RMSE for the 1-st tissue (Squamous)

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE(>0,95)	0,0607	0,0593	0,0499	0,0325
RMSE(PCA)	0,0502	0,0621	0,0507	0,0326

As seen from Table 1 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

**Table 2.** RMSE for the 2-nd tissue (Columnar)

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE (>0,95)	0,0459	0,0457	0,0449	0,0295
RMSE (PCA)	0,0471	0,0480	0,0408	0,0297

As seen from Table 2 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

**Table 3.** MSE for the 3-rd tissue (SEA)

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE (>0,95)	0,0561	0,0530	0,0492	0,0244
RMSE (PCA)	0,0580	0,0577	0,0582	0,0247

As seen from Table 3 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

**Table 4.** MSE for the 4-th tissue (CIN1)

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE (>0,95)	0,0462	0,0852	0,0408	0,0220
RMSE (PCA)	0,0817	0,0958	0,0458	0,0224

As seen from Table 4 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

**Table 5.** MSE for the 4-th tissue (CIN2):

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE (>0,95)	0,0427	0,0497	0,0345	0,0158
RMSE (PCA)	0,0392	0,0500	0,0352	0,0158

As seen from Table 5 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

**Table 6.** MSE for the 4-th tissue (CIN3):

	BP Conjugate gradient	BP Quasi Newton	BP Scaled Conjugate gradient	RBF NN
RMSE (>0,95)	0,0734	0,0789	0,0739	0,0305
RMSE (PCA)	0,0656	0,0755	0,0877	0,0306

As seen from Table 6 RBF network gives the least error, thus both methods of data preprocessing are equivalent.

## Conclusion

- The results of experiments proved that NN are applicable for solving problems described in this paper.
- Experiments showed that RBF network always gives the best results.
- Different methods of data preprocessing improves the NN performance. Both methods proved to be sufficiently effective.
- The developed approach of uterine cervix cancer diagnostics using NN can be recommended for further clinical approbation.

## Acknowledgement

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

## Bibliography

- [Schoonmaker J. 2007] Schoonmaker J. et al. Automatic Segmentation of Uterine Cervix for in vivo Localization and Identification of Cervical Intraepithelial Neoplasia // Apogen Technologies 2007
- [Воробьева Л. И. 2008] Воробьева Л. И. Цитологический скрининг рака шейки матки // Здоровье Украины, 2008
- [Koss L.G. 1989] Koss L.G. The complex test for cervical cancer detection. In: The Journal of the American Medical Association. – 1989
- [Зайченко Ю.П. 2004] Зайченко Ю.П. Основи проектування інтелектуальних систем. Навчальний посібник – К.: Видавничий Дім «Слово», 2004
- [Domagoj Kovacevic, Sven Loncaric 1997] Radial Basis Function-based Image Segmentation using a Receptive Field //Computer-Based Medical Systems. – 1997. Proceedings., Tenth IEEE Symposium , 1997
- [M.C. Jones and R. Sibson 1987] What is projection pursuit ? // J. of the Royal Statistical Society, ser. A – 1987.

## Authors' Information

**Kateryna Malyshevskya** – Ph.d. student, National Technical University of Ukraine "Kyiv Polytechnic Institute", P.O. Box: 03056, Kyiv-56, Ukraine; e-mail: [volovik\\_katya@yahoo.com](mailto:volovik_katya@yahoo.com)

## ИНСТРУМЕНТАЛЬНАЯ СРЕДА ДЛЯ ИССЛЕДОВАНИЯ ЭВОЛЮЦИОННЫХ СТРАТЕГИЙ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ МЕТАМОДЕЛЕЙ

Павел Афонин

**Аннотация:** В статье представлена инструментальная среда для исследования эволюционных стратегий, которые используют механизм аппроксимации целевой функции с помощью аппарата нейронных сетей. Приводится описание алгоритма эволюционной стратегии и подходы к построению метамоделей. Рассмотрены существующие на сегодняшний день алгоритмы оптимизации на основе эволюционных стратегий и метамоделей. Отмечается актуальность применения механизмов адаптации в таких алгоритмах. Описаны основные функции и возможности инструментальной среды. Средством реализации является программный пакет MatLab v.7.1.

**Ключевые слова:** инструментальная среда, эволюционная стратегия, нейронная сеть, метамодел, оптимизация.

**Conference:** The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009

---

### Введение

---

В последние несколько лет эволюционные алгоритмы нашли широкое применение для решения большого числа сложных задач оптимизации, таких как: задачи оптимального проектирования, задачи оптимального управления, задачи оптимизации на основе имитационного моделирования и многих других задач, требующих применения методов глобальной оптимизации. Однако, во многих случаях, требуется большое число расчетов целевой функции для сходимости эволюционного алгоритма. Проблема усугубляется тем, что для многих задач целевая функция представлена неявно, т.е. в виде некоторой расчетной модели, имитационной модели или компьютерной программы, что требует проведения длительных расчетов, которые во многих случаях невозможно выполнить за приемлемое время.

Основным способом решения данной проблемы является использование метамоделей. Метамоделю принято называть приближенную математическую модель, полученную в результате экспериментов с моделью системы с целью замещения последней при оптимизации. Применяются следующие подходы к построению метамоделей: полиномы, кригинг модели и нейронные сети.

Таким образом, актуальной задачей является разработка эффективных эволюционных алгоритмов оптимизации, которые используют метамодел и, как следствие, разработка инструментария для проведения исследования таких алгоритмов.

В статье приводится описание инструментальной среды для исследования эволюционных стратегий, которые используют механизм аппроксимации целевой функции с помощью аппарата нейронных сетей. В качестве эволюционного алгоритма выбрана эволюционная стратегия, потому что этот алгоритм работает достаточно быстро и обладает возможностью адаптации параметров. В качестве метамоделей используются нейронные сети, поскольку они являются мощным аппаратом для аппроксимации сложных зависимостей. В статье дано описание алгоритма эволюционной стратегии, подходы к построению метамоделей и существующие алгоритмы оптимизации на основе эволюционных стратегий и метамоделей. Отмечается актуальность применения механизмов адаптации в таких алгоритмах. Инструментальная среда реализована с помощью программного средства MatLab v.7.1. Представлены итоги работы и планы для будущей работы.

## Описание эволюционной стратегии

Задачей эволюционной стратегии (ЭС) является минимизация целевой функции  $F(x_1, x_2, \dots, x_n)$ , где  $x_i$  – вещественные переменные ( $i=1, 2, \dots, n$ ). В ЭС каждая особь характеризуется:

- функцией пригодности (ФП), которая зависит от целевой функции оптимизационной задачи;
- строкой-хромосомой, включающей:
  - вектор  $x(x_1, x_2, \dots, x_n)$ , который представляет собой некоторое решение оптимизационной задачи;
  - среднеквадратическое отклонение (шаг мутации)  $\sigma(\sigma_1, \sigma_2, \dots, \sigma_m)$ ,  $1 \leq m \leq n$ , от которого зависит величина мутации;
  - угол ротации  $\alpha(\alpha_1, \alpha_2, \dots, \alpha_k)$ ,  $\alpha_j \in [-\pi, \pi]$ ,  $k = n^*(n-1)/2$ , который является необходимым параметром для реализации коррелируемой мутации, позволяющей учитывать ландшафт целевой функции.

В процессе работы алгоритма ЭС происходит адаптация значений шага мутации и угла ротации. Таким образом, происходит двухуровневый процесс оптимизации: на одном уровне осуществляется поиск оптимального значения переменных, на другом – настройка параметров ЭС.

Далее опишем общую схему эволюционной стратегии [3]:

**Шаг 1. Инициализация.** Задается начальная популяция из  $\mu$  особей. В большинстве случаев используется случайная генерация начальной популяции по равномерному закону распределения. Также возможно равномерное задание популяции, задание в области предполагаемого оптимума, задание методом латинских квадратов и др. Значение шага мутации определяется, как правило, разработчиком. Значение угла ротации выбирается случайным образом на интервале  $[-\pi, \pi]$  с помощью равномерного закона распределения.

**Шаг 2. Расчет ФП особей.** Для всех  $\mu$  особей популяции осуществляется расчет ФП. Данный шаг выполняется только в случае применения  $(\mu+\lambda)$ -селекции (см. шаг 6).

**Шаг 3. Скрещивание.** Результатом данного шага является генерация  $\lambda$  потомков из  $\mu$  родителей. Для получения одного потомка сначала происходит случайный выбор двух родителей из популяции с вероятностью  $1/\mu$  для каждого. Затем осуществляется рекомбинация, при которой потомок наследует фенотипические признаки родителей. Рекомбинация реализуется для всей строки-хромосомы, включая шаг мутации и угол ротации.

**Шаг 4. Мутация.** Реализуется мутация  $\lambda$  потомков. Сначала производится изменение параметров стратегии: шага мутации с помощью логнормального закона распределения и угла ротации с помощью нормального закона распределения. После этого реализуется мутация переменных решения на основе случайного вектора, полученного с помощью измененных параметров ЭС:

$$\sigma'_i = \sigma_i \cdot \exp(\tau' \cdot N(0, 1) + \tau \cdot N_i(0, 1));$$

$$\alpha'_j = \alpha_j + \beta \cdot N(0, 1);$$

$$x' = x + N(0, C');$$

где  $N(0, 1)$  – значение случайной величины, полученное с помощью нормального закона распределения с математическим ожиданием 0 и среднеквадратическим отклонением 1;  $\tau'$  – постоянный коэффициент (рекомендуемая величина:  $\tau' \approx 1/(2 \cdot n)^{1/2}$ );  $\tau$  – коэффициент, варьируемый при каждой мутации (рекомендуемая величина:  $\tau \approx 1/(2 \cdot n^{1/2})^{1/2}$ );  $\beta$  – постоянный коэффициент (рекомендуемая величина:  $\beta \approx 5^\circ$ );  $C'$  – ковариационная матрица, полученная после мутации значений  $\sigma$  и  $\alpha$ ;  $N(0, C')$  – случайный вектор, полученный из ковариационной матрицы  $n$ -мерного нормального распределения.

**Шаг 5. Расчет ФП потомков.** Осуществляется расчет ФП для  $\lambda$  потомков текущей популяции.

**Шаг 6. Селекция.** Существует два типа селекции:  $(\mu, \lambda)$ -селекция и  $(\mu+\lambda)$ -селекция. В случае  $(\mu, \lambda)$ -селекции в следующую популяцию выбираются лучшие  $\mu$  родителей (имеющие лучшие значения ФП)

только из множества  $\lambda$  потомков. При  $(\mu+\lambda)$ -селекции новая популяция образуется из объединенного множества родителей и потомков. Более предпочтительной является  $(\mu, \lambda)$ -селекция, поскольку реализует механизм выхода из локальных оптимумов.

**Шаг 7. Проверка условия останова ЭС.** Применяются классические критерии останова для популяционных алгоритмов, такие как: максимальное число итераций (расчетов ФП), максимальное время работы алгоритма, число поколений без изменения лучшего значения ФП, малая разница между лучшим и средним значением ФП и др. Если условие останова не выполняется, то осуществляется переход к шагу 3 для реализации следующего поколения.

---

### Способы построения метамоделей

---

**Полиномы.** Наиболее часто используются полиномы второй степени. Число коэффициентов такой модели рассчитывается как:  $(n+1)*(n+2)/2$ , где  $n$  – число переменных. Для расчета неизвестных коэффициентов полинома можно использовать метод наименьших квадратов или градиентный метод. Основным недостатком метода наименьших квадратов является значительные временные затраты на расчет коэффициентов модели в случае решения задачи большой размерности.

**Кригинг модели.** Представляют собой комбинацию глобальной модели и локальных «отклонений»:

$$y(x) = g(x) + Z(x),$$

где:  $g(x)$  является функцией глобальной модели для целевой функции;  $Z(x)$  представляет собой функцию Гаусса с нулевым математическим ожиданием и ковариацией, моделирующей локальные отклонения от глобальной модели. Обычно, функция  $g(x)$  задается полиномом или, во многих случаях, задается как коэффициент  $\beta$ .

Расчет параметров модели осуществляется с помощью метода максимального правдоподобия.

Основным достоинством кригинг моделей является возможность расчета доверительного интервала без дополнительных вычислений. Однако необходимо выполнять матричные преобразования для расчета выхода модели, что значительно увеличивает время вычислений с ростом размерности задачи.

**Нейронные сети.** Являются мощным аппаратом для аппроксимации сложных зависимостей [4]. Здесь применяются три типа сетей: многослойный персептрон, сети на основе радиальных базисных функций и машины опорных векторов.

Многослойный персептрон прямого распространения состоит из: входного слоя, состоящего из множества сенсорных элементов; одного или нескольких скрытых слоев вычислительных нейронов (как правило, с сигмоидальной функцией активации) и выходного слоя нейронов. Обучение такой сети выполняется с помощью алгоритма обратного распространения ошибки (BP-алгоритм). Для повышения эффективности решения задачи с помощью многослойного персептрона применяют модификации BP-алгоритма и методы оптимизации структуры сети для конкретной задачи.

Сеть на основе радиальных базисных функций (RBF-сеть) состоит из трех слоев. Входной слой включает сенсорные элементы, которые связывают сеть с внешней средой. Промежуточный слой является единственным скрытым слоем сети. Он состоит из радиальных элементов, каждый из которых воспроизводит гауссову поверхность отклика и выполняет нелинейное преобразование входного пространства. Для большинства задач размерность скрытого слоя значительно превышает размерность входного слоя. Выходной слой RBF-сети включает нейроны с линейными функциями активации и выполняет линейное преобразование пространства скрытого слоя нейронов.

Машина опорных векторов представляет собой линейную систему, которая решает задачу разделения объектов в пространстве признаков с помощью гиперплоскости. Теория таких систем исходит из теории статистического обучения. Основными достоинствами машины опорных векторов являются: отсутствие локального минимума ошибки в процессе обучения, а также то, что ошибка обобщения не зависит от размерности задачи.

---

## Эволюционные алгоритмы оптимизации на основе метамоделей

---

**Интеграция метамоделей в эволюционные алгоритмы.** Существует два основных подхода к интеграции метамоделей в эволюционные алгоритмы. В первом подходе сначала определяется оптимум по метамоделю, а затем реализуется расчет оптимизируемой функции в точке оптимума. Новые данные используются для построения новой (более точной) метамоделю и процесс определения оптимума и обновления метамоделю повторяется. Второй подход основан на концепции *эволюционного контроля* [5], в рамках которого применяются два метода: эволюционный контроль на уровне особей и эволюционный контроль на уровне поколений. В методе эволюционного контроля на уровне особей часть особей текущей популяции рассчитывается с помощью целевой функции. Для остальных особей в популяции расчет ФП реализуется с использованием метамоделю. В методе эволюционного контроля на уровне поколений все особи некоторой популяции рассчитываются или с помощью целевой функции или с помощью метамоделю. Особи, ФП которых рассчитывается с помощью целевой функции, называются *контролируемые*, а для которых с помощью метамоделю – *неконтролируемые*.

**Эволюционный контроль на уровне особей.** В настоящее время данный подход представляется интересным и малоисследованным [1, 2, 5, 6]. Здесь главной задачей является определение того, какие особи в каждом поколении будут рассчитываться с помощью целевой функции, а какие с помощью метамоделю. Опишем два базовых метода эволюционного контроля на уровне особей: метод на основе оценки лучших особей (*best* стратегия) и метод с применением предварительного отбора (*pre-selection* стратегия).

В *best* стратегии [5],  $\lambda' = \lambda$  потомков рассчитываются с помощью метамоделю и лучшие  $\lambda^*$  рассчитываются с помощью ФП. После обновления метамоделю оставшиеся  $\lambda' - \lambda^*$  потомков снова рассчитываются по метамоделю. Лучшие  $\mu$  особей из  $\lambda$  потомков становятся родителями следующего поколения.

В *pre-selection* стратегии [6],  $\lambda' > \lambda$  потомков генерируются из  $\mu$  родителей с помощью операторов рекомбинации и мутации и затем рассчитываются с помощью метамоделю. Лучшие  $\lambda^* = \lambda$  особей предварительно отбираются из  $\lambda'$  потомков и рассчитываются с помощью ФП.

По результатам исследований, описанных в работе [2], сделано важное заключение о том, что стабильность эволюционной стратегии на основе эволюционного контроля на уровне особей может быть улучшена, если родители для следующего поколения выбираются из контролируемых особей, подобно тому, как это реализовано в *pre-selection* стратегии.

Однако на сегодняшний день остается открытым вопрос о том, сколько особей и какие особи текущего поколения должны контролироваться.

**Механизмы адаптации.** Основным механизмом адаптации является следующее правило: если качество метамоделю *улучшается* в процессе поиска, то *большее* число особей должно рассчитываться с использованием метамоделю.

В работе [6] предложено правило определения числа особей для предварительного отбора  $\lambda_{pre}$  в зависимости от качества селекции, рассчитываемого по метамоделю. В работе [2] описаны механизм адаптации на основе расчета среднеквадратического отклонения, механизм адаптации на основе селекции и механизм адаптации на основе корреляции.

Существует точка зрения [2, 6], что при использовании метамоделю в эволюционных алгоритмах важным является только правильная селекция, а не ошибка аппроксимации метамоделю.

---

## Описание инструментальной среды

---

На рисунке 1 представлено диалоговое окно инструментальной среды для исследования эволюционных стратегий с использованием нейросетевых метамоделю. Текущая версия поддерживает алгоритмы, реализующие метод эволюционного контроля на уровне особей. Среда разработана в среде MatLab 7.1, программный код функций и алгоритмов содержится в 5-ти m-файлах.



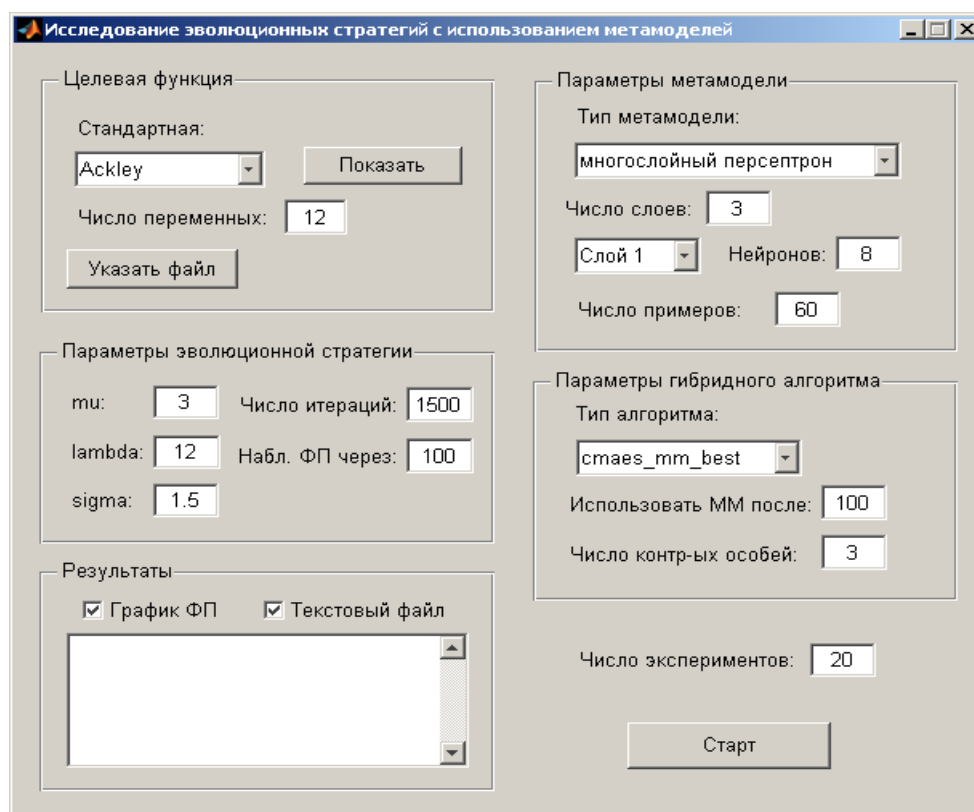


Рис.1. Диалоговое окно инструментальной среды

Опишем возможности и соответствующие блоки инструментальной среды:

**Целевая функция.** Задаются параметры оптимизируемой функции. Можно выбрать стандартную тестовую функцию (вкладка *стандартная*) или загрузить из файла (указать М-файл, в котором содержится определение функции). Предусмотрены стандартные тестовые функции: *Sphere*, *Schwefel*, *Ackley*, *Rosenbrock*, *Rastrigin* и др. Имеется возможность просмотра в отдельном окне ландшафта (поверхности) стандартных тестовых функций в трехмерной системе координат. В файле можно задать определение как простой функции (функция задается аналитически), так и сложной (функция задается как компьютерная программа, работающая по определенному алгоритму). Также можно задать размерность исследуемой функции (параметр *число переменных*).

**Параметры эволюционной стратегии.** Задаются параметры алгоритма эволюционной стратегии: тип стратегии ( $\mu, \lambda$ ) и шаг мутации  $\sigma$ , а также число итераций (расчетов ЦФ) работы алгоритма и интервал, с которым осуществляется наблюдение значений ФП в процессе работы алгоритма, необходимый для сбора статистических данных по результатам отдельных экспериментов.

**Параметры метамодели.** Задаются параметры нейросетевой метамодели. Во вкладке *тип метамодели* можно выбрать тип нейронной сети, на основе которой будет реализовано построение метамодели: многослойный перцептрон, сеть на основе радиальных базисных функций, машина опорных векторов. Можно задавать параметры нейронной сети и число примеров обучения. Так, например, для многослойного перцептрона можно задать число скрытых слоев и число нейронов в каждом скрытом слое.

**Параметры гибридного алгоритма.** Текущая версия поддерживает следующие алгоритмы: *rigestmaes* (базовый алгоритм эволюционной стратегии без использования метамоделей), *pre-selection* стратегия, *best* стратегия и вновь разработанный алгоритм «*my\_cmaesmm*», в котором число контролируемых особей для одного поколения может изменяться в зависимости от точности модели, а выбор  $\mu$  родителей для следующего поколения всегда осуществляется из контролируемых особей. В будущем, планируется опубликовать результаты исследований эффективности данного алгоритма по сравнению с другими

стратегиями. Также исследователю предоставляется возможность задавать число контролируемых особей. Имеется возможность задать число примеров для обучения НС после которого начинается построение метамоделей.

**Результаты.** Имеется возможность проведения серии экспериментов с заданными параметрами, а также расчет средних значений по результатам серии экспериментов. Планируется реализация расчета доверительных интервалов и среднеквадратических ошибок. Результаты можно представить в виде текстового файла, в котором описаны средние и лучшие значения ФП для каждого эксперимента (имеется возможность задать число экспериментов) через определенное число итераций. В отдельном блоке приводятся результаты работы алгоритмов, которые можно использовать для окончательной оценки их эффективности. Результаты по каждому эксперименту сохраняются в массиве данных, который можно использовать в дальнейшем.

---

### Заключение

Статья посвящена разработке инструментальной среды для исследования эволюционных стратегий, которые используют механизм аппроксимации целевой функции с помощью аппарата нейронных сетей. В качестве эволюционного алгоритма выбрана эволюционная стратегия, а в качестве метамоделей используются нейронные сети. Рассмотрены эволюционные алгоритмы оптимизации на основе метамоделей. Инструментальная среда реализована с помощью программного средства MatLab v.7.1.

В дальнейшем планируется расширение среды и добавление следующих возможностей: включение алгоритмов, реализующих метод эволюционного контроля на уровне поколений, добавление других способов построения метамоделей (полиномы и кригинг модели), включение других эволюционных алгоритмов, в частности, различных схем реализации генетического алгоритма.

---

### Библиография

1. M. Emmerich, A. Giotis, M. Özdenir, T. Bäck, and K. Giannakoglou. Metamodel-assisted evolution strategies. In *Parallel Problem Solving from Nature*, number 2439 of *Lecture Notes in Computer Science*, pp. 362-370, 2002.
2. L. Gräning, Y. Jin, and B. Sendhoff. Individual-based Management of Meta-models for Evolutionary Optimization with Application to Three-Dimensional Blade Optimization. *Evolutionary Computation in Dynamic and Uncertain Environments*, pp. 225-250, 2007.
3. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2), pp. 159–196, 2001.
4. S. Haykin. *Neural Networks – A Comprehensive Foundation*. Prentice-Hall, 1994.
5. Y. Jin, M. Olhofer, and B. Sendhoff. A framework for evolutionary optimization with approximate fitness functions. *IEEE Transactions on Evolutionary Computation*, 6(5), pp. 481-494, 2002.
6. H. Ulmer, F. Streichert, and A. Zell. Evolution strategies with controlled model assistance. In *Congress on Evolutionary Computation*, pp. 1569–1576, 2004.
7. П. Афонин. Система оптимизации на основе имитационного моделирования, генетического алгоритма и нейросетевых метамоделей // *Proceedings of the XIII-th International Conference «Knowledge-Dialogue-Solutions» – Varna, 2007. – V.1. – P.60-63.*

---

### Информация об авторе

**Павел Афонин** – к.т.н.; Московский Государственный Технический Университет им. Баумана; 2-я Бауманская, д.5, Москва, Россия; e-mail: [pavlafon@yandex.ru](mailto:pavlafon@yandex.ru)

## ПРИМЕНЕНИЕ НЕЙРОННОЙ СЕТИ ХЕММИНГА И НЕЧЕТКОЙ ЛОГИКИ К ОБНАРУЖЕНИЮ КРАЕВ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ В ОТТЕНКАХ СЕРОГО

Николай Мурга

**Аннотация:** В данной работе исследуется применение нейронной сети Хемминга для обнаружения краёв объектов на изображении. Изображение в оттенках серого, поступающее на вход предлагаемой системы, подвергается преобразованию с применением нечёткой логики в двуцветное. После этого из изображения последовательно выделяются блоки пикселей заданной размерности и подаются на входы предварительно инициализированной сети Хемминга. Нейронная сеть выполняет идентификацию краёв в блоке, и в новом изображении вставляет на место блока шаблон, который отвечает коду, полученному на выходе сети. Работу завершает практическое применение метода.

**Ключевые слова:** Детектирование краёв объектов изображения, Метод разностного группирования, Нейронная сеть Хемминга, Нечёткая логика.

**ACM Classification Keywords:** I.4.3. Enhancement – Grayscale manipulation, I.4.6. Segmentation – Edge and feature detection, I.4.6. Segmentation – Pixel classification.

**Conference:** The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009

---

### Введение

В процессе медицинской диагностики часто возникает необходимость сегментации изображений. Медицинские изображения, а в частности рентген-изображения, характеризуются сильной зашумленностью. Для преодоления этого негативного фактора вышеуказанные методы должны обладать низкой чувствительностью к шумам. Одним из подходов к разрешению данной проблемы является применение нейронных сетей. В данной работе предлагается метод, который использует нечёткую логику и нейронную сеть Хемминга для обнаружения краёв объектов на рентген-изображениях. Изображение, попадающее на вход предлагаемой системы, подвергается преобразованию в двуцветное, путём применения нечёткой логики. Прежде всего, пиксели изображения подвергаются кластеризации методом разностного группирования по признаку цвета. Выделяются центры кластеров и из них выбираются наибольший и наименьший (однако при модификациях метода возможно использование и большего числа центров). Функции принадлежности для кластеров выбираются гауссовского вида. Определяются среднеквадратические отклонения. Так строится нечёткое правило, позволяющее принять решение о том, к какой конкретно цвет имеет данный пиксель. Решение принимается на основе максимального значения принадлежности кластеру среди кластеров. После этого применяется нейронная сеть Хемминга, принцип функционирования которой описан в первом разделе статьи. По изображению перемещается окно. В данной работе для практических примеров окно выбрано размером  $4 \times 4$ . Данные окна пикселей преобразовываются таким образом, чтобы их можно было подать на входы сети Хемминга. Сеть Хемминга выдаёт номер образца, который наиболее похож на поданный на входы. На основе полученного номера из базы шаблонов выбирается соответствующий шаблон и записывается на место в новом, перед этим созданным, изображении, которое (место в новом изображении) соответствует месту в изображении поданного на вход сети примера. После «прохода» всех окон получается новое изображения

с детектированными краями, но не цельными. Работу завершает применение метода на сегментах рентген изображений с сильными шумами.

### Постановка задачи

На входе имеется изображение в оттенках серого.

Необходимо:

1. Преобразовать данное изображение в изображение, которое будет иметь два отличающихся цвета.
2. На основе полученного в пункте 1 изображения, применяя сеть Хемминга, произвести идентификацию краёв объектов на изображении.

### Нейронная сеть Хемминга

Описание структуры и алгоритма функционирования данной нейронной сети предлагается в работах [1], [2], [3], [4]. В текущем разделе даётся описание структуры и функционирования сети Хемминга, объединяющее информацию из вышеуказанных трудов.

Сеть состоит из двух слоёв. Первый и второй слои имеют по  $m$  нейронов, где  $m$  - это число примеров. Нейроны первого слоя имеют по  $n$  синапсов, соединённых с входами сети (которые образуют фиктивный нулевой слой). Нейроны второго слоя связаны между собой ингибиторными (взаимно обратными) синаптическими связями. Единственный синапс с позитивной обратной связью для каждого нейрона соединён с его же аксоном.

Идея работы сети состоит в поиске расстояния Хемминга от примера, который тестируется, ко всем примерам.

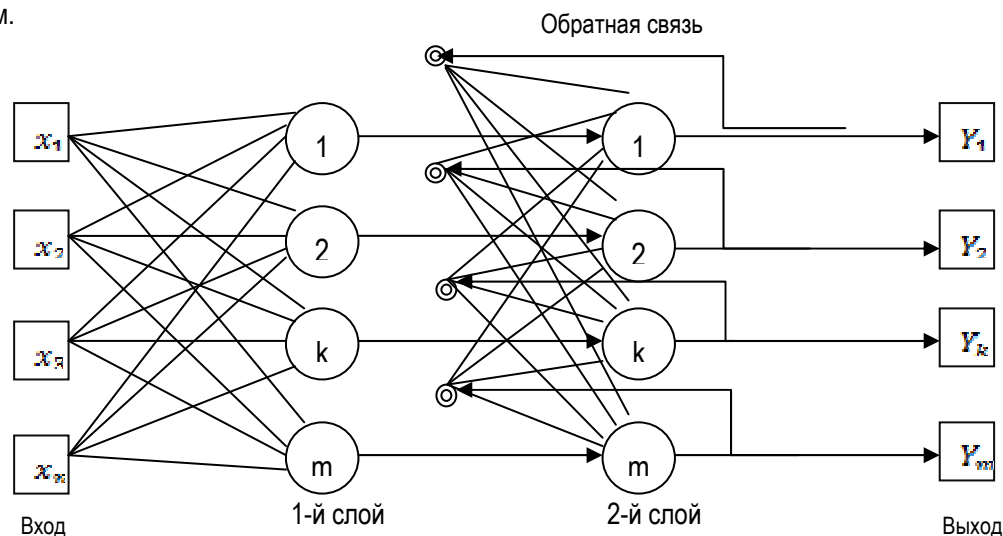


Рис. 1 Структурная схема сети Хемминга

Расстоянием Хемминга называется число отличных битов в двух бинарных векторах. Сеть должна выбрать образец с минимальным расстоянием Хемминга от неизвестного входного сигнала, в результате чего будет активирован только один выход сети, который отвечает этому образцу.

На стадии инициализации весовых коэффициентов первого слоя и порога активационной функции им будут присвоены следующие значения:

$$\omega_{ik} = \frac{x_i^k}{2}, i = 0 \dots n - 1, k = 0 \dots m - 1, \quad (1)$$

$$T_k = \frac{n}{2}, k = 0 \dots m-1, \quad (2)$$

где  $x_i^k$  –  $i$ -й элемент -го примера.

Весовые коэффициенты тормозящих синапсов во втором слое принимаются равными величине,  $-\varepsilon$  где  $\varepsilon \in \left[0, \frac{1}{m}\right]$ . Синапс нейрона, связанный с его же аксоном, имеет вес +1.

Алгоритм функционирования сети Хемминга следующий.

*Шаг 1.* На входы сети подаётся неизвестный вектор, исходя из которого рассчитываются состояния нейронов первого слоя (верхний индекс в скобках над переменной обозначает номер слоя)

$$y_j^{(1)} = f(s_j^{(1)}) = f\left(\sum_{i=0}^{n-1} \omega_{ij} x_i + T_j\right), j = 0 \dots m-1. \quad (3)$$

После этого, полученными значениями инициализируются значения аксонов второго слоя.

$$y_j^{(2)} = y_j^{(1)}, j = 0 \dots m-1. \quad (4)$$

*Шаг 2.* Вычислить новые значения входов нейронов второго слоя

$$s_j^{(2)}(p+1) = y_j^{(2)}(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), k \neq j, j = 0 \dots m-1 \quad (5)$$

и значения их аксонов

$$y_j^{(2)}(p+1) = f(s_j^{(2)}(p+1)), j = 0 \dots m-1. \quad (6)$$

Активационная функция  $f$  имеет вид порога, при этом величина порога должна быть достаточно большой, чтобы любые возможные значения аргумента не приводили к насыщению.

*Шаг 3.* Проверяется, изменились ли входы нейронов второго слоя за последнюю итерацию. Если да – перейти к шагу 2. Иначе – конец.

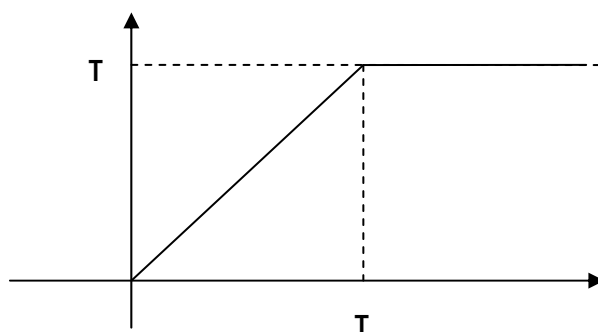


Рис. 2 Вид пороговой функции активации. T следует брать достаточно большим

### Метод разностного группирования

Далее описывается метод разностного группирования, взятый из работы [2]. Алгоритм разностного группирования – это модификация алгоритма пикового группирования, в котором векторы, подлежащие кластеризации  $x_j$ , рассматриваются как потенциальные центры кластеров. Пиковая функция  $D(x_i)$  задаётся формулой:

$$D(x_i) = \sum_{j=1}^N \exp \left\{ - \frac{\|x_i - x_j\|^{2b}}{\left(\frac{r_a}{2}\right)^2} \right\}, \quad (7)$$

где значение коэффициента  $r_a$  определяет сферу соседства. На значения  $D(x_i)$  значительно влияют только  $x_j$ , которые находятся в пределах данной сферы.

При большой плотности точек вокруг  $x_i$  значение функции  $D(x_i)$  большое. После расчёта значений пиковой функции для каждой точки  $x_i$ , выбирается вектор  $x$ , для которого мера плотности  $D(x)$  окажется самой большой. Именно эта точка и становится первым центром  $c_1$ .

Выбор следующего центра  $c_2$  возможен после исключения предыдущего центра и всех точек, лежащих в его окрестности.

Пиковая функция переопределяется следующим образом:

$$D_{new}(x_i) = D(x_i) - D(c_1) \cdot \exp \left\{ - \frac{\|x_i - c_1\|^{2b}}{\left(\frac{r_b}{2}\right)^2} \right\}. \quad (8)$$

При новом определении функции  $D$  коэффициенты  $r_b$  обозначают новые значения константы, которая задаёт сферу соседства очередного центра. Обычно придерживаются условия, что  $r_b \geq r_a$ .

После модификации значения пиковой функции ищется новая точка  $x$ , для которой  $D_{new}(x) \rightarrow \max$ . Она становится новым центром.

Процесс поиска очередного центра возобновляется после исключения всех компонент, которые отвечают уже отобраным точкам. Инициализация завершается в момент фиксации всех центров, которые предусмотрены начальными условиями.

В соответствии с описанным алгоритмом происходит самоорганизация множества векторов  $x$ , которая состоит в нахождении оптимальных значений центров, которые представляют множество данных с минимальной погрешностью.

---

### Предлагаемый метод и его практическое применение

---

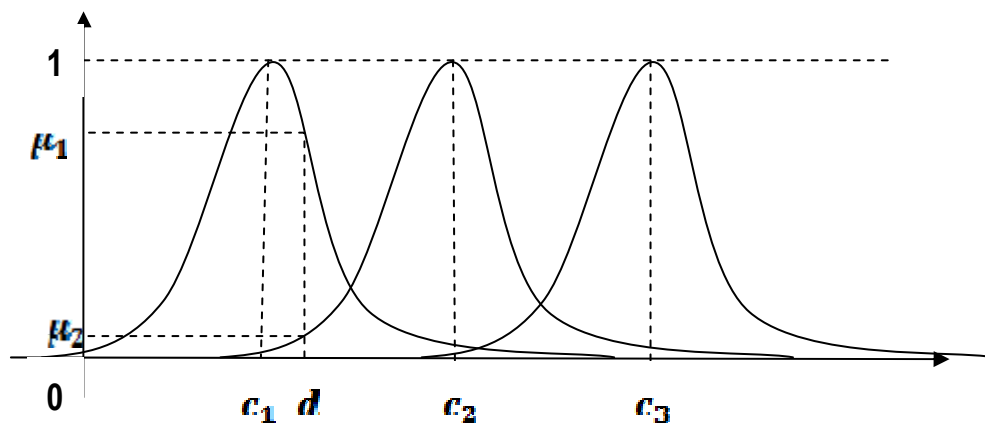
Суть метода состоит в следующем. Если из изображения выделять изображения, которые являются частью первого (в дальнейшем в работе такие изображения будут называться подизображениями), то количество видов краёв подизображения будет, в силу физических обстоятельств (количества пикселей), меньше разнообразия краёв исходного изображения. Если взять размерность подизображения достаточно малой (например,  $3 \times 3$ ,  $4 \times 4$ ,  $5 \times 5$  и т.п.), то количество видов краёв (границ) в изображении можно без труда пересчитать. Очевидно, что каждый из указанных только что краёв можно сопоставить некоторой букве некоторого алфавита. Таким образом, каждый вариант можно закодировать определённым символом.

Из всего вышеуказанного можно сделать следующий вывод. Всё изображение можно разбить на непересекающиеся подизображения, каждое из которого принимает значение из некоторого алфавита. То есть, на вход системы подаётся вектор, соответствующий подизображению, а на выходе система должна предлагать своё видение, какому символу из алфавита данный вектор соответствует. Подобные условия совпадают с требованиями к функционированию сети Хемминга, что и обосновывает её применение для решения данной проблемы.

Далее предлагается метод уменьшения входного алфавита для системы. Край – это линия, отделяющая один цвет от другого. Таким образом, для системы выделения краёв нет необходимости учитывать все оттенки цветов, подлежащих разделению. Системе необходимо сказать к какому из цветовых классов пикселей принадлежит данный. Для разрешения этой проблемы предлагается следующий метод.

Всё изображение поддается кластеризации цветов методом разностного группирования. Это позволяет выбрать центры кластеров, к которым «притягиваются» остальные цвета. Если выбрать функции принадлежности к кластеру гауссовскими и выбрать определённым образом дисперсию для данного закона, то получается система, изображённая на рисунке 3.

На данном рисунке  $c_1, c_2, c_3$  - это центры кластеров в одномерном пространстве оттенков серого.  $d$  - это цвет, который подлежит классификации.  $\mu_1, \mu_2$  и  $\mu_3$ , которая в данном случае близка к нулю, - это значения принадлежности  $d$  соответственно к каждому из классов. Если  $i = \overline{1,3}$  - это номер класса, то номер класса, к которому, скорее всего, принадлежит  $d$  - это  $i_d = \arg \max_i \mu_i$ . Т.о., подавая на вход данной системы цветное значение пикселя, на выходе получается номер класса, к которому принадлежит пиксель.



**Рис. 3 Нечёткая система определения цвета**

Алгоритм идентификации краёв, который предлагается, состоит в следующем.

*Шаг 0.* Происходит инициализация сети Хемминга, её памяти. Строится база шаблонов, в которой каждый шаблон отвечает определённому выходу сети Хемминга. На вход системы подаётся изображение в оттенках серого. Создаётся новое пустое изображение, которое совпадает по размерам с изображением, которое поступило на вход системы.

*Шаг 1.* Каждый пиксель поступившего изображения проходит через систему классификации цвета, которая определяет его цвет. Для метода необходимо, чтобы изображение было преобразовано к двуцветному виду. Следовательно, выбирается для классификации два крайних («слева» и «справа») класса.

*Шаг 2.* Из поступившего изображения выделяется подизображение малого размера (например,  $3 \times 3$ ,  $4 \times 4$ ,  $5 \times 5$  и т.п.).

*Шаг 3.* Определяется, есть ли на подизображении значительное соотношение пикселей различных цветов. Если да – та шаг 4. В противном случае, изображение делается одноцветным и – переход на шаг 6.

*Шаг 4.* Подизображение подаётся на вход сети Хемминга, которая определяет, какому шаблону оно более всего соответствует.

*Шаг 5.* Имея полученный на выходе сети Хемминга код, система отыскивает в базе шаблонов для заполнения подходящий шаблон и ставит его на место в новом изображении, которое соответствует месту подизображения в старом изображении, которое подавалось на вход сети Хемминга.

*Шаг 6.* Проверка того, есть ли необработанные подизображения на старом изображении. Если есть – на шаг 2. В противном случае – на шаг 7.

*Шаг 7.* Выход системы – полученное новое изображение.

Далее предлагаются примеры практического применения предлагаемой системы. В качестве объектов применения системы выбраны сильно зашумленные части рентген изображений.

Подизображения выбираются размером  $4 \times 4$ . Они не пересекаются. Первоначальное изображение «пропускается» через систему классификации цвета пикселей, которая описана выше и здесь рассматривается только два класса цветов. Таким образом, входная буква алфавита сети – 16 бит, а выходной алфавит выбран в 18 символов.

В качестве порога, когда подизображение «пропускать» через сеть Хемминга не имеет смысла, взяты величины 0,125 и 0,875 (для данного случая). Если  $s$  - число пикселей в подизображении, которые имеют один из цветов, а  $n$  – общее число пикселей в подизображении, то для того, чтобы подизображение подверглось классификации при помощи сети Хемминга, необходимо, чтобы  $0,125 < \frac{s}{n} < 0,875$ . То

есть, для рассматриваемого случая, если в подизображении количество пикселей одного из цветов меньше 2, то подизображение не подаётся на вход сети Хемминга и просто заполняется одним цветом.

Следующие примеры сопровождаются также приведением результатов работы детектора Канни ([5]) на тех же изображениях.



Рис. 4 Исходное изображение

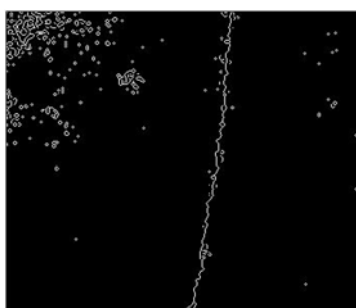


Рис. 5 Изображение после применения детектора Канни



Рис. 6 Изображение после применения предлагаемого метода



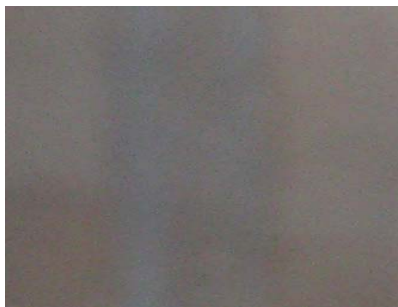


Рис. 7 Исходное изображение

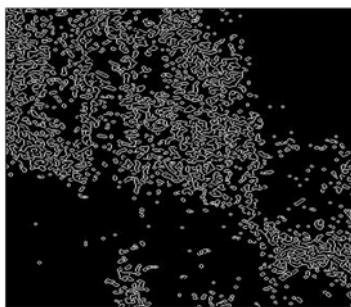


Рис. 8 Изображение после применения детектора Канни



Рис. 9 Изображение после применения предлагаемого метода

Следует сразу же сказать несколько слов о моменте, который бросается в глаза. Предлагаемый метод не соединял края – не делал их цельными. Но это и не было целью данной работы. Для соединения краёв можно применить любой из существующих методов (например, на основе динамически изменяющегося радиуса окрестности пикселя, находить в данной окрестности наиболее близкий, совпадающий по цвету и отдалённый, пиксель и соединять его с центром окрестности (обрабатываемым в данный момент пикселем)). Однако, не смотря на это, очевидно, что предлагаемый метод правильно распознаёт края, хотя он и несколько чувствителен к шуму. Но чувствительность эта оказалась ниже, чем у детектора Канни, хотя детектор Канни делал края более цельными. Результаты обработки изображения, показанного на рисунке 4 – это рисунки 5 и 6. А на рисунке 7 – 8 и 9. Из данных рисунков видно, что в первом случае оба метода справились с работой довольно хорошо. Второй же случай – даже человеком не всегда может быть достаточно хорошо обработан, что и обуславливает зашумленность результатов работы методов. Однако в обоих случаях предлагаемый метод выдал меньшее число шумов, чем детектор Канни. Скорость работы методов была практически одинаковая.

Безусловно, как показывают вышеприведённые примеры, детектор Канни делал края практически замкнутыми, что нельзя сказать про предлагаемый метод. Этот очень важный факт обуславливает необходимость дальнейшего совершенствования предлагаемого метода. Прежде всего, необходимо расширить порог для определения необходимости «пропускания» подизображения через сеть Хемминга путём расширения размерности подизображения. Однако, очевидно, что расширение размерности должно происходить в разумных пределах, так как с расширением размерности растёт и входной и выходной алфавиты. Но это может быть и позитивной стороной, так как растёт разнообразие случаев, которые может обрабатывать метод. Кроме того, необходимо провести исследование о возможности сокращения букв выходного алфавита сети Хемминга, что может стать причиной более качественного распознавания поступившего символа сетью Хемминга и меньшей чувствительности к шумам. То есть разнообразие символов может и не быть позитивной стороной. Так же не обязательно подизображения не должны пересекаться, так как изучение именно их пересечения может дать дополнительную информацию для определения краёв. И следует ещё отметить, что изображение для обработки должно быть не обязательно двуцветным, однако в случае большего числа цветов необходимы дополнительные надстройки, позволяющие корректно функционировать сети Хемминга.

---

## Заключение

---

В работе был предложен метод идентификации краёв изображений на основе применения сети Хемминга. Практическое применение метода позволило сделать ряд выводов. Первый, метод действительно идентифицирует края. Второй, метод менее чувствителен к шуму чем, например, детектор Канни. Третье,

скорость работы методов примерно одинакова. Однако, несмотря на это, метод значительно проигрывает детектору Канни по цельности обнаруженных краёв. Это обуславливает необходимость дальнейшего совершенствования предлагаемого детектора. Во-первых, очевидно, что рассмотренный размер окна не является оптимальным и должны быть проведены исследования по поиску оптимального размера окна, так как простое увеличение одновременно ведёт к росту алфавита, используемого для обработки данных. Во-вторых, необходимо изучение возможностей «скользящего окна», то есть возможности устранения ограничения на не пересечение выделяемых блоков пикселей. Главные проблемы здесь – как объединять данные про один и тот же пиксель (здесь необходимо использовать возможности теории Демпстера-Шэффера, например) и в то же время минимизировать чувствительность к шумам. Все эти исследования будут проведены в последующих работах.

---

### Благодарности

---

Статья частично финансированна из проекта **ITHEA XXI** Института Информационных теории и Приложений FOI ITHEA и Консорциума FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

### Литература

---

1. Борисов Е.С. Классификатор на основе нейронной сети Хемминга – 2005 – взято с <http://mechanoid.ru/nns/classifiers/>
2. Зайченко Ю.П. Основы проектирования интеллектуальных систем.- К.: Видавничий Дім «Слово», 2004. – 352 с.
3. Короткий С. Нейронные сети Хопфилда и Хемминга – взято с [http://lii.newmail.ru/kor\\_nn4.htm](http://lii.newmail.ru/kor_nn4.htm)
4. Стариков А., Арустамов А. Ассоциативная память – применение сетей Хемминга для нечёткого поиска. – <http://www.basegroup.ru/library/analysis/neural/assoc/>
5. Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6): pages 679-698, November 1986.

---

### Информация про автора

---

**Мурга Николай Алексеевич** – магистрант НТУУ «КПИ» УНК «ИПСА», Киев, Украина;  
e-mail: [murqa.nicholas@gmail.com](mailto:murqa.nicholas@gmail.com)

## ОТОБРАЖЕНИЕ И ВЫВОД ПО АНАЛОГИИ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ РАСПРЕДЕЛЕННЫХ ПРЕДСТАВЛЕНИЙ

Сергей Слипченко, Дмитрий Рачковский

**Аннотация:** *Развит подход к рассуждениям по аналогии для иерархически структурированных описаний эпизодов, ситуаций и их компонентов на базе представлений аналогов в виде особой формы векторных представлений - распределенных кодвекторных представлений. Предложены распределенные представления компонентов аналогов, позволяющие непосредственно определять соответствующие друг другу представления компонентов для реализации стадии отображения двух аналогов, а также метод вывода по аналогии на их основе. Предложенные методы исследованы на базах аналогий, которые ранее применялись для исследования ведущих моделей аналогии - SME и ACME. Полученные результаты находятся на уровне результатах SME и ACME, однако за счет использования сходства векторных представления обладают низкой вычислительной сложностью и создают основу для более адекватного учета семантики аналогов и их компонентов. Это делает предложенные методы перспективными для отображения фрагментов баз знаний с большим числом компонентов.*

**Ключевые слова:** *аналогия, отображение аналогов, вывод по аналогии, распределенное представление информации, кодвекторы, базы знаний, SME, ACME*

**ACM Classification Keywords:** *1.2 Artificial Intelligence, 1.2.4 Knowledge Representation Formalisms and Methods, 1.2.6 Learning (Analogies)*

**Conference:** *The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS 2009, Varna, Bulgaria, June-July 2009*

---

### Введение

---

Продуктивным подходом, используемым людьми при решении сложных задач в условиях неполноты, неточности, противоречивости входной информации, являются рассуждения на основе примеров решения похожих задач. Для многих предметных областей и задач существенным является использование реляционной структурированной информации, где в явном виде присутствуют описания объектов и отношений разных уровней иерархии. Например, в базах знаний (БЗ), онтологиях, и др. Рассуждения по аналогии – это вид рассуждений на примерах, где существенное внимание уделяется сходству реляционных структур, то есть систем отношений между объектами, в ситуациях, предметных областях, эпизодах и т.п., которые могут выглядеть непохожими при поверхностном рассмотрении [Gentner, 1983], [Hummel & Holyoak, 1997].

Мышление с использованием аналогий является одним из важнейших процессов разумной деятельности людей и его моделированию посвящено большое количество работ [Gentner, 1983], [Holyoak & Thagard, 1989], [Hummel & Holyoak, 1997], [Markman, 1997], [Гладун, 2000], [Гладун и др.]. В рассуждениях по аналогии первыми тремя стадиями считают [Falkenhainer et al., 1989], [Holyoak & Thagard, 1989], [Hummel & Holyoak, 1997] поиск (процесс обнаружения в памяти наиболее близкого аналога ко входному), отображение (установление соответствия между компонентами двух аналогов) и вывод по аналогии (процесс переноса знаний от одного аналога к другому). Все эти стадии требуют обработки структурированной информации, содержащейся в представлении аналогов.

Аналоги есть иерархически структурированные описания эпизодов или ситуаций в БЗ. Реализация и эффективность операций над структурированной информацией (оценка сходства, сравнение, нахождение соответствующих элементов, обход, и др.) существенно зависит от ее представления.

Для оценки сходства структурированной информации иногда привлекают подходы, использующие понятие частичного изоморфизма [Falkenhainer et.al., 1989], [Holyoak & Thagard, 1989]. Его выявление для графов является вычислительно сложной (NP) задачей. Однако изоморфизм не учитывает сходство самих компонентов и не отражает особенностей оценки сходства, обнаруженных психологами в рассуждениях по аналогии у людей. В связи с важностью моделирования рассуждений по аналогии для проблематики искусственного интеллекта (ИИ) актуальной является разработка подходов и методов, которые позволяли бы достигать уровня результатов лучших известных (символьных) моделей [Falkenhainer et.al., 1989], [Holyoak & Thagard, 1989], однако позволяли бы преодолеть их недостатки [Eliasmith & Thagard, 2001], [Hummel & Holyoak, 1997], [Kanerva, 2000], [Plate, 2003] (высокая вычислительная сложность и слабый учет семантического сходства компонентов аналогов).

На базе бинарных нейросетевых распределенных представлений (РП) структурированной информации, разрабатываемых в рамках парадигмы ассоциативно-проективных нейронных сетей, авторами развит новый подход к моделированию рассуждений по аналогии, направленный на применение рассуждений по примерам в системах ИИ, основанных на знаниях [Markman et.al., 2003], [Rachkovskij, 2004], [Rachkovskij & Kussul, 2001], [Рачковский и Слипченко, 2005]. В данной статье рассматриваются методы отображения, основанные на бинарных разреженных РП иерархических реляционных структур, оценивается их вычислительная сложность и качество функционирования.

### Распределенные представления отношений

Нейросетевой подход привел к идее РП информации – форме векторного представления, где каждый объект (признак, физический объект, их совокупность, отношение, сцена и др.) представлен совокупностью элементов вектора, а отдельный элемент вектора может принадлежать представлениям разных объектов. Авторами исследуются методы РП информации, которые трансформируют описания объектов  $x$  в кодвекторы  $\mathbf{X}$  ( $x \rightarrow \mathbf{X}$ ). Кодвектор – это форма векторного представления информации со свойствами бинарности ( $\mathbf{X} \in \{0,1\}^M$ ) и разреженности (доля ненулевых элементов  $M$  кодвектора  $\mathbf{X}$  размерности  $N$  мала:  $M/N \ll 1$ ). Сходные (по векторным мерам сходства) кодвекторы должны соответствовать сходным в контексте решаемой задачи объектам.

Реляционная структурированная информация современных декларативных баз знаний (БЗ) рассматривается в работе как помеченный направленный упорядоченный ациклический граф, где вершины-потомки (аргументы отношений, представленных вершинами-родителями), в свою очередь, могут быть как объектами, так и отношениями (рис. 1).

Разработаны методы формирования кодвекторов отношений вида  $R(A,B,\dots)$ , (где  $R$  – идентификатор отношения,  $A,B,\dots$  – аргументы), соответствующие схемам *роль-заполнитель* (*role-filler*) и *предикат-аргументы* (*predicate-arguments*), традиционно используемым в символьных представлениях. Методы основаны на использовании процедур связывания кодвекторов [Rachkovskij & Kussul, 2001] (функционального аналога группирующих

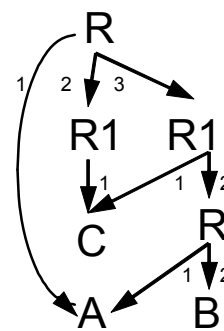


Рис. 1. Пример фрагмента БЗ

скобок для представления совокупности компонентов в символьной нотации):  $(a,b,c\dots) \rightarrow \langle \mathbf{A}, \mathbf{B}, \mathbf{C}, \dots \rangle$ , где  $(a,b,c\dots)$  – компоненты группы,  $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$  – их кодвекторы,  $\langle \dots \rangle$  – процедура связывания кодвекторов.

Методом роль-заполнитель кодвектор отношения формируется как  $R(A,B,\dots) \rightarrow \langle \langle \mathbf{R}_a, \mathbf{A} \rangle, \langle \mathbf{R}_b, \mathbf{B} \rangle, \dots \rangle$ , где  $\mathbf{A}, \mathbf{B}, \dots$  – кодвекторы аргументов (объектов-заполнителей),  $\mathbf{R}_a, \mathbf{R}_b, \dots$  – кодвекторы ролей.

Методом предикат-аргументы кодвектор отношения формируется как  $R(A,B,\dots) \rightarrow \langle \mathbf{R}_{-1}, \mathbf{A}_{-2}, \mathbf{B}_{-3}, \dots \rangle$ , где  $\mathbf{R}, \mathbf{A}, \mathbf{B}, \dots$  – кодвекторы предиката (отношения) и аргументов,  $\mathbf{R}_{-i}$  – связывание перестановкой кодвектора, где  $i$  – порядковый номер аргумента (задающий его роль).

Размерность кодвекторов отношений и их аргументов одинакова, что позволяет формировать кодвекторы сложных иерархических реляционных структур, содержащих отношения высших порядков, рекурсивным применением разработанных методов построения кодвекторов отношений (роль-заполнитель и предикат-аргументы). Например, для реляционной структуры рис. 1 кодвектор, сформированный методом предикат-аргументы, имеет вид:

$$\langle \mathbf{R} \vee \mathbf{A}_{-1} \vee \langle \mathbf{R}_1 \vee \mathbf{C}_{-1} \rangle_{-2} \vee \langle \mathbf{R}_1 \vee \mathbf{C}_{-1} \vee \langle \mathbf{R} \vee \mathbf{A}_{-1} \vee \mathbf{B}_{-2} \rangle_{-2} \rangle_{-3} \rangle. \tag{1}$$

Для структур со сходными объектами и отношениями продуцируются сходные кодвекторы. Сходство кодвекторов тем больше, чем больше сходство отношений и их аргументов.

### Отображение аналогов с помощью распределенных представлений отношений

Разработаны методы отображения, основанные на использовании сходства кодвекторов, которые впервые позволили реализовать отображение аналогов со сложной структурой с помощью РП.

Поскольку аналогия основана на реляционном сходстве, при отображении существенным является не только сходство компонентов (атрибутов и отношений, входящих в данный компонент), но и сходство их ролей в отношениях более высокого уровня (в какие атрибуты и отношения входит данный компонент). Для учета этого разработан метод формирования промежуточных представлений компонентов аналогов согласно подходу, предложенному в [Rachkovskij, 2004], [Рачковский и Слипченко, 2005]. Метод заключается в объединении векторных представлений компонентов с соответствующими им ролями.

**Солнце\*** = *масса(Солнце) ∨ больше(масса(Солнце), масса(Планета)) ∨ гравитация(Солнце, Планета) ∨ притягивает(Солнце, Планета) ∨ ...*

Представление  $\mathbf{V}^*$  для рассматриваемого компонента  $v$  определяется как дизъюнкция представлений компонента  $\mathbf{V}$  и представлений всех ролей для всех возможных путей из вершины  $v$ :

$$\mathbf{V}^* = \mathbf{V} \vee \bigvee_{w \in W_l(v)} \mathbf{V}_w, \tag{2}$$

где  $\mathbf{V}_w$  – вектор, соответствующий роли  $w$ ,  $W_l(v)$  – множество возможных дуг ролей вершины  $v$ ,  $l(v)$  – уровень вершины  $v$ .

$W_l(v)$  определяется следующим образом:

$$W_0(v) = \{(v, v') \in E\};$$

$$W_{i+1}(v) = W_i(v) \cup \{(v', v'') \in E : (v, v') \in W_i(v)\}.$$

Уровень вершины  $l(v)$  вычисляется как:

$$l(v) = \begin{cases} 1 + \max_{(v,u) \in E} l(u), & \text{если } \exists u (v,u) \in E \\ 0, & \text{иначе} \end{cases}$$

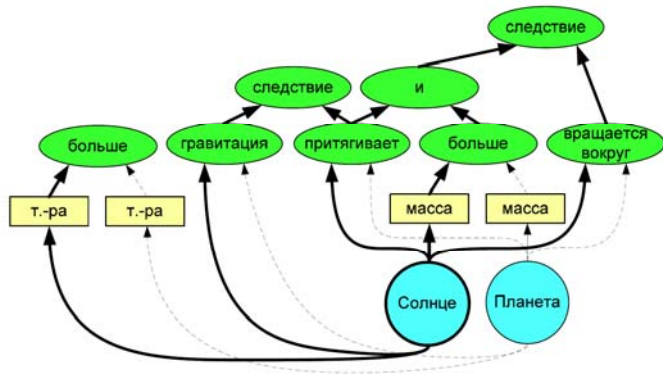


Рис. 2. Объединение ролей объекта Солнце

где  $E$  – множество вершин графа  $G$ ;  $u$  – вершина, инцидентная выбранной вершине  $v$  (в терминах аналогов это означает, что высказывание  $v$  является аргументом предиката высказывания  $u$ ).

Пример формирования промежуточного представления для отображения объекта Солнце приведен на рис. 2.

Отображение осуществляется следующим образом:

1. Для высказываний базового аналога уровня 0 находятся наилучшие отображения среди компонентов целевого аналога по максимуму скалярного произведения их векторных представлений  $\mu(x) = \arg \max_{\forall y} (\mathbf{V}_x, \mathbf{V}_y)$ , где  $x$  принадлежит  $\{u \in V_S: I(u) = 0\}$  – подмножеству графа вершин базового

аналога  $G_S$  с уровнем 0, а  $y$  принадлежит  $V_T$  – множеству всех вершин графа целевого аналога  $G_T$ .

2. Учитывая ограничение структурной согласованности (аргументы отношений поставленных в соответствие, также должны соответствовать), выполняется рекурсивное сопоставление аргументов высказываний в глубь по дереву аргументов, которые были отображены на шаге (1). Для каждой пары сопоставляемых аргументов вычисляется скалярное произведение их промежуточных представлений, полученных по формуле (2). Пары, у которых скалярное произведение промежуточных представлений меньше случайного перекрытия, исключаются. В результате формируется список троек  $(x, y, q)$ , где  $x, y$  – сопоставленные компоненты;  $q$  – критерий качества (значение скалярного произведения их промежуточных представлений).

3. Наилучшее отображение компонента  $x$  базового аналога на компонент целевого аналога  $y^*$  находится по максимуму суммы критериев качества для данного отображения, т.е.  $y^* = \arg \max_{\forall y} \sum_{(x, y, q_i): x_i = x, y_i = y} q_i$

Предложенный алгоритм отображения позволяет учитывать как сходство самих компонент, поскольку оно отражается в сходстве соответствующих им векторных представлений, так и сходство ролей отображаемых компонент в отношениях и атрибутах, содержащих данные компоненты, за счет формирования промежуточных представлений.

Результаты предложенных методов для сложных аналогов совпадают с результатами психологических тестов и результатами лучших известных систем *SME* и *ACME*.

Вычислительная сложность предложенного метода отображения  $O(n^2M)$ , где  $n$  – среднее число элементов аналогов, а  $M$  – среднее число единиц в векторах представлений.

---

### Вывод по аналогии с помощью распределенных представлений отношений

---

Известные модели вывода по аналогии основаны на копировании с подстановкой и дополнением [Markman 1997], где выполняется копирование отношений из базового аналога в целевой с подстановкой соответствующих (если есть), или созданием новых компонент аналогов (дополнение).

На рис. 3 показан процесс переноса отсутствующей причинно-следственной связи между притяжением ядра и электрона, и вращением электрона вокруг ядра. Предположение сделано на основе сходства моделей атома и солнечной системы, где существует связь между притяжением Солнца и Планеты, и вращением Планеты вокруг Солнца.

Важную роль при построении выводов играет ограничение систематичности – предпочтение глубоких иерархий связанных отношений. В результате на целевой эпизод переносятся только те отношения базового эпизода, которые имеют значительное пересечение с общими компонентами. Общими считаются компоненты базового аналога, которые имеют соответствия в целевом. В примере на рис. 3 разность температур Солнца и Планеты оказалась несущественной для модели атома.



Использование распределенных представлений позволяет естественным образом учитывать свойство систематичности за счет их формирования таким образом, что скалярное произведение векторов возможных новых отношений с вектором целевого аналога оказывается большим, чем пересечение с векторами существующих в целевом аналоге компонентов. В качестве представлений элементов аналогов для вывода используются промежуточные представления, получаемые при формировании вектора аналога для отображения.

Предлагаемый алгоритм вывода по аналогии использует в качестве основы алгоритм поиска наилучших отображений.

1. Для высказываний базового аналога уровня 0 находятся наилучшие отображения среди высказываний целевого аналога по максимуму скалярного произведения их векторных представлений  $\mu(x) = \arg \max_{\forall y} (\mathbf{V}_x, \mathbf{V}_y)$ , где  $x$  принадлежит

$\{u \in V_S: I(u) = 0\}$  – подмножеству графа вершин базового аналога  $G_S$  с уровнем 0, а  $y$  принадлежит  $V_T$  – множеству всех вершин графа целевого аналога  $G_T$ .

2. Для найденных предварительных отображений выполняется поиск наилучших отображений их аргументов. Если величина скалярного произведения векторных представлений отображенных высказываний базового и целевого аналогов  $(\mathbf{V}_x, \mathbf{V}_y)$  меньше, чем величина скалярного произведения отображения любого из его аргументов  $(\mathbf{V}_{x(i)}, \cdot)$ , то данное высказывание принимается в качестве гипотезы о выводе, иначе данное высказывание считается существующим в целевом аналоге. Для аргументов высказываний-гипотез повторяется шаг (2), а все оставшиеся высказывания базового аналога (имеющие отображение в целевом аналоге) обрабатываются на шаге (3).

3. Учитывая ограничение структурной согласованности (аргументы отношений, поставленных в соответствие, также должны соответствовать), выполняется рекурсивное сопоставление аргументов высказываний, которые были отображены на шаге (1). Для каждой пары сопоставляемых аргументов вычисляется скалярное произведение их промежуточных представлений полученных по формуле (2), пары у которых скалярное произведение промежуточных представлений меньше случайного перекрытия исключаются. В результате формируется список троек  $(x, y, q)$ , где  $x, y$  – сопоставленные элементы;  $q$  – критерий качества (значение скалярного произведения их промежуточных представлений).

4. Наилучшее отображение компонента  $x$  базового аналога  $\mu(x)$  находится по максимуму суммы критериев качества для данного отображения, т.е.  $\mu(x) = \arg \max_{\forall y} \sum_{(x_i, y_i, q_i): x_i = x, y_i = y} q_i$

Гипотезы – отношения и атрибуты базового аналога, переносимые на целевой, имеют значительное число компонент общих для обоих аналогов. Это обусловлено тем, что векторные представления гипотез имеют значительное пересечение с векторными представлениями своих, а соответственно и сходных с ними компонент. В сочетании с механизмами отображения, включенными в этот алгоритм, он позволяет

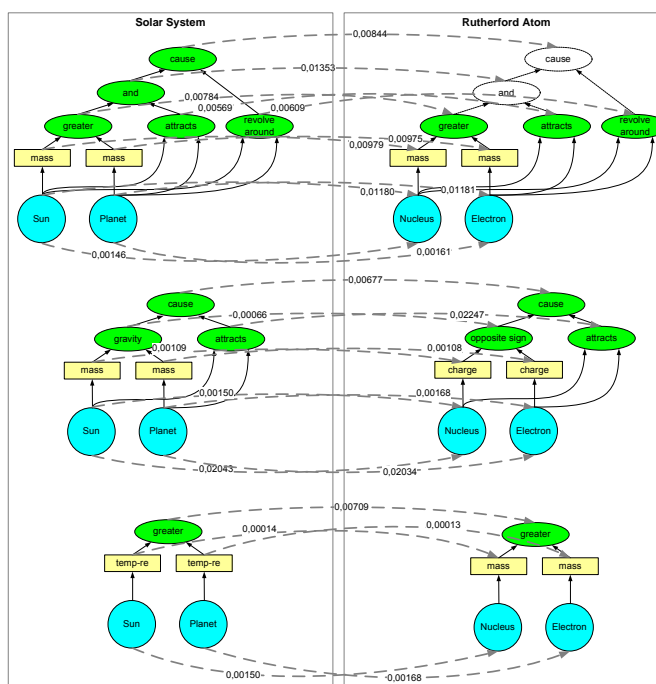


Рис. 3. Перенос выводов из аналога «Солнечная система» (слева) в аналог «Атом Резерфорда» (справа)

---

одновременно решать задачу отображения и вывода по аналогии, формируя в результате список гипотез и отображений упорядоченный по степени «правдоподобия».

Вычислительная сложность процедуры вывода сравнима со сложностью процедуры отображения.

Разработанные методы реализованы программными средствами и экспериментально исследованы на фрагментах БЗ, используемых для исследования моделей рассуждений по аналогии. Проведенные исследования показали адекватность предложенного подхода к моделированию рассуждений по аналогии и более высокую вычислительную эффективность его реализации по сравнению с традиционными символьными методами.

---

## Обсуждение

Анализ моделей аналогии, разработанных на базе символьных и локальных нейросетевых представлений (SME, ACME и др.), показывает, что, хотя такие модели хорошо приспособлены для обработки и сравнения сложных иерархических структур, необходимых для рассуждения по аналогии, оценка сходства аналогов требует отображения их компонентов, реализуемых вычислительно сложными процедурами. На сегодняшний день наиболее развитой теорией аналогии является теория структурного отображения Structure Mapping Theory (SMT) школы D. Gentner [Gentner, 1983], [Falkenhainer et.al., 1989], [Markman, 1997]. В SMT впервые в явной форме делается акцент на структурном сходстве между аналогами, что позволяет строить их отображение независимо от предметной области. На базе SMT построен ряд моделей отображения аналогий: Structure Mapping Engine (SME) [Falkenhainer et.al., 1989] и др. Вычислительная сложность SME составляет от  $O(n^2)$  до  $O(n!)$ . Сложность для модели отображения ACME [Holyoak & Thagard, 1989] составляет  $O(n^4)$ .

Для более адекватного учета семантики аналогов, масштабирования подходов на случаи, когда имеется большое количество потенциальных аналогов, повышения нейробиологической релевантности моделей в новых моделях рассуждений по аналогии стали использовать РП. Однако анализ моделей рассуждения по аналогии LISA [Hummel & Holyoak, 1997] и DRAMA [Eliasmith & Thagard, 2001] показывает, что использование РП в них носит фрагментарный, непоследовательный характер, либо их авторам удалось работать лишь с простейшими аналогами.

При исследовании разработанных кодвекторных методов отображения на базах знаний, на которых ранее исследовались лучшие известные символьные методы SME и ACME, получены результаты, соответствующие уровню их результатов. Соответствие результатов экспериментальных исследований разработанных моделей результатам известных психологических экспериментов и лучших известных моделей подтвердили адекватность предложенного подхода и методов рассуждения по аналогии. Подход также открывает возможности для учета семантического сходства объектов и отношений в моделях рассуждений по аналогии благодаря использованию кодвекторов, сходство которых отражает их семантическую близость. Вычислительная сложность отображения предложенными методами составляет  $O(n \cdot n')$  -  $O(n^2)$ , где  $n'$  – число компонентов, для которых надо найти отображение,  $n$  – число компонентов другого аналога, с которыми надо установить отображение. Это существенно ниже сложности традиционных методов  $O(n^4)$ - $O(n!)$ , что делает предложенные методы перспективными для отображения фрагментов БЗ с большим числом компонентов.

---

## Благодарности

Статья частично финансирована из проекта ITHEA XXI Института Информационных теории и Приложений FOI ITHEA и Консорциума FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).



---

## Литература

---

- [Eliasmith & Thagard, 2001] C. Eliasmith & P. Thagard (2001). Integrating Structure and Meaning: A Distributed Model of Analogical Mapping. *Cognitive Science*, 25(2), 245-286.
- [Falkenhainer et al., 1989] B. Falkenhainer, K.D. Forbus, & D. Gentner (1989) The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41, 1-63.
- [Gentner, 1983] D. Gentner (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, pp. 155-170.
- [Holyoak & Thagard, 1989] K.J. Holyoak, & P. Thagard (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.
- [Hummel & Holyoak, 1997] J.E. Hummel & K.J Holyoak (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, pp. 427-466.
- [Kanerva, 2000] P. Kanerva (2000) Large patterns make great symbols: An example of learning from example. In S. Wermter and R. Sun (eds.), *HYBRID NEURAL SYSTEMS* (pp. 194-203). Heidelberg: Springer.
- [Markman, 1997] A.B. Markman (1997). Constraints on Analogical Inference. *Cognitive Science*, 21(4), pp. 373-418.
- [Markman et al., 2003] A.B. Markman, D.A. Rachkovskij, I.S. Misuno, E.G. Revunova (2003). Analogical reasoning techniques in intelligent counterterrorism systems // *Informational Theories & Applications*, 10(2), pp. 139-146.
- [Plate, 2003] T.A. Plate (2003). *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. Stanford: CSLI Publications, 300 p.
- [Rachkovskij, 2004] D.A. Rachkovskij (2004). Some Approaches to Analogical Mapping with Structure Sensitive Distributed Representations // *Journal of Experimental and Theoretical Artificial Intelligence*, 16(3), pp. 125-145.
- [Rachkovskij & Kussul, 2001] D.A. Rachkovskij, E.M. Kussul (2001). Binding and Normalization of Binary Sparse Distributed Representations by Context-Dependent Thinning. *Neural Computation*, 13(2), pp. 411-452.
- [Гладун, 2000] В.П. Гладун (2000). Партнерство с компьютером: Человеко-машинные целеустремленные системы. К.: Port-Royal, 128 с.
- [Гладун и др., 2000] В.П. Гладун, В.Ю. Величко, Н.Н. Киселева, Н.М. Москалькова (2000). Вывод гипотез о составе и свойствах объектов на основе аналогии. *Искусственный интеллект*, № 1, С. 44-52.
- [Рачковский и Слипченко, 2005] Д.А. Рачковский, С.В. Слипченко (2005). Подходы к отображению аналогов с помощью распределенных представлений. *Компьютерная математика*, №1, С. 55-69.

---

## Информация об авторах

---

**Сергей Слипченко** – научный сотрудник; Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, пр. Акад. Глушкова 40, Киев-03680, Украина;  
e-mail: [serge.slipchenko.irtcits@gmail.com](mailto:serge.slipchenko.irtcits@gmail.com)

**Дмитрий Рачковский** – ведущий научный сотрудник; Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, пр. Акад. Глушкова 40, Киев-03680, Украина; e-mail: [dar@infrm.kiev.ua](mailto:dar@infrm.kiev.ua)



---

---

## Evolutionary and Genetic Algorithms

---

---

### THE PRODUCTION SCHEDULING IN ASSEMBLY SYSTEM WITH EVOLUTIONARY ALGORITHM

**Galina Setlak**

**Abstract:** *In this paper an evolutionary algorithm is proposed for solving the problem of production scheduling in assembly system. The aim of the paper is to investigate a possibility of the application of evolutionary algorithms in the assembly system of a normally functioning enterprise producing household appliances to make the production graphic schedule.*

**Keywords:** *Artificial intelligence, flexible assembly systems, evolutionary algorithm, production scheduling.*

**ACM Classification Keywords:** *I. Computing methodologies I.1.Symbolic and algebraic manipulation I.1.3.Evaluation strategies I.2.Artificial Intelligence I.2.8.Problem solving Control Methods and Search – Scheduling J.6.Computer Aided Engineering - Computer Aided Manufacturing (CAM).*

**Conference:** *The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009*

---

#### Introduction

---

The competition in the world market requires that machine and technological equipment manufactures try new effective tools to maximally shorten the preparation time to start production and meet quality standards. The unusually intensive changes in the market result in the need to realize a growing number of production objectives in the shortest time possible. All this makes the automatization of the assembly process more and more decisive for the competitiveness of a modern enterprise [Sawik, 1996].

The problem of how to optimize planning methods, including production scheduling, has drawn the attention of quite a number of scientists since as early as mid-fifties [Conway, 1967]. The complexity of assembly scheduling and its highly significant effect on the functioning of assembly systems and their production impact necessitate looking for and developing advanced methods and algorithms for solving scheduling related problems. In recent years many research centers have been using artificial intelligence methods including evolutionary algorithms to find out how to optimize production plans.

The aim of the paper is to investigate a possibility of the application of evolutionary algorithms in the assembly hall of a normally functioning enterprise producing household appliances to make the production graphic schedule.

Production scheduling in an assembly system is connected with taking particular decisions about the passage of objects that are being assembled through the system. Here, two basic kinds of scheduling tasks can be distinguished [Sawik, 1996]:

- Initial scheduling of products, which consists in timing the introduction of the successive base parts of different types of assembled products into the system. Decisions on the choice of the product for assembly and the moment when it can leave the factory are taken with regard to operational and tactical plans (assembly hall capacity, preplanned technological routes).

- Scheduling assembly and transport operations involving the assigning of assembly and time limits to particular machines. That kind of scheduling is curbed by the inflow of the particular types of base parts to the system, i.e. the results of the assumed initial schedule.

Thus, the two basic schedules above are closely connected. While making detailed production schedules it is common practice to aim at attaining a specific objective which is later used as an evaluation criterion for the schedule. The chosen targets may be, e.g. time minimization of all the tasks, reduction of delays in meeting buyers' expectations, making full use of the production potential and others.

---

### What are evolutionary algorithms

---

The basic idea of evolution modeling has always been intriguing for researchers in many fields of science: "Let us replace the process of modeling human race by modeling its evolution".

Evolutionary algorithms are computer-aided problem solution systems. They are based upon the principles that can be observed in the evolution of living organisms [Goldberg, 1989]. The idea of evolutionary algorithms is founded on the processes observed in nature such as the selection of specimens and evolution of species, reproduction mechanism and inheriting characteristics.

Evolutionary algorithms include also such methods as genetic algorithms, evolutionary programming and evolutionary strategies [Goldberg, 1989], [Michalewicz, 1996]. The paper, following the recent trends, makes use of the general term and generally accepted name: evolutionary algorithms (EA). Because of the size limits of the paper the differences between the methods have not been explained here.

The working of evolutionary algorithms can be described in a simple, step-by-step way as follows:

1. Initiation-the formation of the initial population of specimens, which means a random choice of the necessary number of chromosomes (specimens). In such a population each specimen represents an acceptable solution.
2. Evolution of chromosome adaptation in the population-involves calculating the value of the adaptation function (fp) for each chromosome of a particular population. The value of (fp) depends on what kind of problem is being solved.
3. Checking if the termination criterion has been met-that depends upon a particular application of EA. If the criterion has been fulfilled, we pass on to the final step which is extracting the the best chromosome. If it does not happen this way, the next step is selection.
4. Chromosome selection, which consists in choosing the chromosomes which will participate in the creation of offsprings for the new generation. The process follows the principles of natural selection. The chromosomes with the greatest fp value have the best chance to participate in the creation of new offsprings. There are many selection methods [Goldberg, 1989], [Pawlak, 1999], but the simplest and the most popular one is the roulette method as its randomness is like that of a roulette. The probability of picking a particular chromosome is greater with an increase in the fp value. The selection results in creating a parent population whose size equals that of the current population.
5. Application of genetic operators to the selected chromosomes, which leads to the creation of a new population which is made up of the populations of offsprings obtained from the selected population of their parents. In EA two basic genetic operators are applied:
  - crossover,
  - mutation operator.

The chromosomes of the chosen parents are combined to produce offsprings. The process is also called recombination.

Mutation is done after crossover and involves the introduction of some random alterations in the chains of descendant chromosomes created before. Like in nature, mutation occurs extremely rarely: the probability of its occurrence of is very small ( $0 \leq p_m \leq 0.01$ ).

6. Evaluation of all the offsprings that form the population.

Not only the newly obtained descendant chromosomes but also all the others are evaluated.

7. Creation of a new population.

At all times the offsprings that are the fittest stand the best chance of getting into the new population. Here various methods can be applied [Michalewicz, 1996], [Pawlak, 1999].

Now we pass on to step 3.

Steps 4-7 are performed loop wise until the criterion of EA termination has been met.

8. Derivation of the best chromosome.

If the EA has been stopped, the effect of the algorithm should be derived, i.e. the solution to the problem should be found. The best solution is the chromosome of the greatest fp value.

---

### **An example of the evolution algorithm application to assembly process scheduling**

---

The chapter will present the results of the practical application of evolution algorithms to scheduling the production at the vacuum cleaner assembly department of the ZELMER Household Appliances Factory. Currently the factory is producing seven basic models of vacuum cleaners. Their technological designs are different, but the number of modules and the number of the parts that make them up is basically the same. A manual and the assembly instruction materials for the vacuum cleaner which was probably the prototype of one of currently produced models were used for the research.

The following assumptions are made in order to work out the production schedule:

- Work schedule applies to the completion of monthly tasks of the assembly department in the two-shift system.
- Nine tasks are realized in the production process and each of them means putting together a defined number of modules making up a finished product as well as the final assembly of a fixed number of one particular model.

As sub-assembly and complete product assembly operations are regarded as tasks, there are some sequence restrictions between them. Fig 1. shows the basic components singled out in the assembly and disassembly operations of a vacuum cleaner.

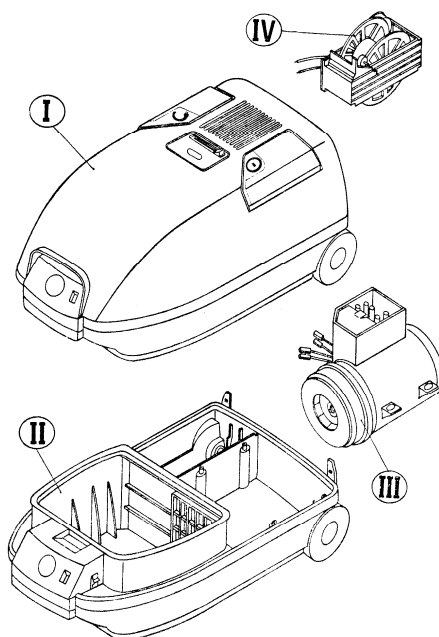


Fig 1. Basic assembly components of a vacuum cleaner [Instruction, 1998].

A finished vacuum cleaner (I) consists of 4 basic subassemblies

- II-complete body consisting of 18 parts-components,
- III-a sucking set composed of 9 elements,
- IV-a complete power cord reel made up of 18 parts,
- Apart from II-IV sets, a complete vacuum cleaner includes 24-26 other parts.

One of the most essential decisions while applying evolutionary modeling methods is specifying the space of the solutions to be searched by the evolutionary algorithm. This is attained through defining the mapping between a point in the solutions space (schedule) and a point in the representation space, i.e. chromosome. There are two known approaches to solving schedule problems with evolutionary algorithms which apply two kinds of representation [Pawlak, 1999]:

- Direct representation-involves using the schedule as the chromosome. Here, the schedule for a particular machine is an arranged list of time limits of starting the operations performed with the machine. The method needs special crossovers to guarantee that the sequence of the operations necessary to carry out the task is not disturbed.
- Indirect representation evolves using the sequence of the tasks to be done as the chromosome. At the chromosome level the tasks have no fixed technological plans, reserves or the beginning and conclusion times. The chromosome is made acceptable by means of a special decoder (schedule making procedures). In the assembly system example under consideration, due to the particular task completion sequence, a simplified variant of the approach presented in [Setlak, 2004] was used where the chromosome is represented by the sequence:

$$(Z_1, s_1) (Z_2, s_2) \dots (Z_n, s_n)$$

where  $Z_i$  - i-task,  $s_i$ -obligatory arrangement following the sequence of performing i-task.

Thus, sequence limitations between tasks are coded in the chromosome. Fig.2 shows the hierarchical structure of the finished product, which basically determines the sequence of the tasks to be performed. This way, there is also a possibility to code the assembly priorities of particular vacuum cleaner types to meet the buyer's expectations.

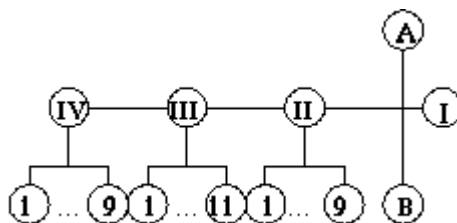


Fig.2. Structure of finished product, arranged to follow the sequence of assembly operations.

In the structure in Fig.2, 1-9 or 1-11 mark the operations performed while putting together a particular sub-assembly described above as A - assembly of sucking pipes involving 4 operations and B - assembly of a small suction nozzle involving one operation. The assembly of the whole vacuum cleaner, marked as I, involves 19 operations that are not shown in the picture.

The assembly of the particular types of the vacuum cleaners, which slightly differ in fittings or subassemblies, follows the same sequence, as shown in Fig.2. In the case of only one type the order is different as it involves one more subassembly to perform one more function. This has been considered by placing one element (sentence) in a chromosome with another number of sub-assemblies and operations.

In the schedule creation process, i.e. arranging the particular elements of the chromosome, a principle was adopted that all the next task operations are fixed in the schedule at the earliest possible dates of their commencement. Thus, assembly operations of sub-assembly IV come first-from IV.1 to IV.9. Then, successively, such operations of sub-assembly III, from III.1 to III.11, which is in agreement with the assembly order of the

finished product given in manual [Instruction, 1998]. The next chromosome element are all the II 1-II 9 fitting operations of sub-assembly II, to be finally followed by operations I1-I. 19, A.1-A.4 and B1 to arrive at the finished product. Consequently, the chromosome will be represented by sequence:

$$(Z_1, IV_1, 1, IV_2, 2, \dots, IV_9, 9, III_1, 1, 10, III_2, 11, \dots, III_11, 21, II_1, 22, II_2, 23, \dots, II_9, 30, I_1, 31, \dots, I_19, 50, A_1, 51, \dots, A_4, 54, B, 55) (Z_2, IV_2-B_2) \dots (Z_9, IV-B_9).$$

For the sake of simplicity detailed operations are not included in the chromosome representation and the exemplifying chromosome is written as:

$$(Z_7, 1)(Z_5, 2)(Z_2, 3)(Z_9, 4)(Z_3, 5)(Z_4, 6)(Z_1, 7)(Z_8, 8) (Z_6, 9)$$

In respect of the defined production quota and the two-shift working system (48 shifts per month) it was necessary to divide the production tasks into parties, which increased the number of components in the chromosome up to 37 elements.

The structure of the evolutionary algorithm can be described as follows:

**Step 1.** The creation of the initial population was carried out by the heuristic method described in [Conway, 1967].

**Step 2.** Creation of the acceptable chromosome, due to restrictions.

**Step 3.** Chromosome evaluation. While working on the presented system, one of the basic evaluation criteria, i.e. minimization of the time necessary to finish all the tasks ( $C_{max}$ ) was made use of. Because of this there is a need to transform the evaluation into the adaptation function which will be maximized. Let us rewrite the adaptation function following the formula:

$$F(x) = \frac{\max_x [C_{max}(x)] - C_{max}(x) + \gamma}{\max_x [C_{max}(x)] - \min_x C_{max}(x) + \gamma} \quad (1),$$

where  $F(x)$ -value of adaptation function,

$\max_x [C_{max}(x)]$  - maximum importance of completion data in the particular generation,

$C_{max}(x)$ -value of completion date for individual  $x$ ,

$\min_x [C_{max}(x)]$  - minimum time limit value in the particular generation,

$\gamma$ - coefficient which can perform two functions.

Depending on the situation occurring during the solution of the problem it can take small values from interval (0,1) so that the quotient denominator does not equal zero. In the other case, coefficient  $\gamma$  is used for scaling congruence function and leveling the differences between individuals in the population. In such a case the value of  $\gamma$  must be correspondingly greater and it is a parameter of the system [Mattfeld, 1996].

**Step 4.** Memorizing the best chromosome.

**Step 5.** Chromosome selection was made with a roulette wheel and by the tournament method.

**Step 6.** Crossover. For that purpose a special method was worked out. It is based on the PMX (Partially Mapped Crossover) method presented in [Uckun, 1993]. The modification involves the introduction of the element of searching extra solution space.

**Step 7.** Mutation. It did not always occur while solving a problem posed.

**Step 8.** Repairing effect of chromosomes.

**Step 9.** Putting the best chromosome in the population.

**Step 10.** Evaluation of chromosomes.

**Step 11.** Memorizing the best chromosome.

In the algorithm steps 1-4 are initial ones and made once only, whereas steps 5-11 are performed many times until the condition of evolution termination has been met.

## Results of calculation experiments

A few series of test were made while testing the evolution algorithm for scheduling assembly tasks that is described in the paper. The first test series was performed at the following assumptions:

- population size-50
- crossover by the modified PMX method.
- application of order-based mutation
- chromosome evaluation made with help of adaptation function described by formula (1), the analyzed various at different  $\gamma$  coefficient values:  $\gamma=0$ ,  $\gamma=0.1$
- In each evolution process 500 generations were considered.
- Application of evolution strategy by choosing the best chromosome and putting it in the next generation to replace the randomly chosen one.
- Simulations for various crossover and mutation probability values were carried out. Each evolution process was repeated ten times and the average result was calculated.

To analyze and estimate the schedules obtained by applying the worked-out evolution algorithm, a heuristic algorithm of active schedule and no-delay schedule were created.

In the building process of no-delay schedules the priority rules considered were as follows:

- the shortest operation, i.e. from among the available alternative assembly operations we select the one whose completion time is the shortest (SPT rule [Conway, 1967])
- the shortest completion time limit rule.

Table 1. shows the results obtained by the methods discussed in the paper.

Table 1. The results obtained by the methods

Method	Evolutionary Algorithm	No-delay schedule	Active schedule
$C_{max}(x)$ (h)	368,2	412,86	434,12
Calculation time	1,45 h	14 sek	15 sek

## Conclusion

The results obtained by the evolutionary algorithm method are much better than those by heuristic scheduling or the no-delay method. However, we must notice the relatively long calculation time by the evolutionary algorithm which is 1.45 hours.

Based on the tests carried out, whose results have only partially been presented in the paper, it can be pointed out that evolutionary algorithms are a very effective tool that enables solving complicated practical optimization problems including NP-hard production scheduling problems in assembly systems. An important characteristic of evolutionary algorithms is their simplicity and versatility. Their main drawback is a long calculation time, which, however, is not a serious disadvantage nowadays with advanced computer technology and does not limit their use for searching for almost optimal solutions.



---

**Bibliography**

---

- [Conway, 1967] Conway R.W., Maxwell W.L., Miller L.W. Theory of scheduling. Addison-Wesley, Reading, Massachusetts, 1967, 382 str.
- [Goldberg, 1989] Goldberg D.E. Genetic algorithms in search. Optimization and Machine Learning, Addison-Wesley P.C., 1989.
- [Mattfeld, 1996] Mattfeld D.C. Evolutionary search and the job-shop. Physica - Verlag, Heidelberg, 1996.
- [Michalewicz, 1996] Michalewicz Z. Algorytmy genetyczne + struktury danych = programy ewolucyjne, WNT, Warszawa, 1996, 432 str..
- [Pawlak, 1999] Pawlak M.: Algorytmy ewolucyjne jako narzędzie harmonogramowania produkcji, PWN, Warszawa, 1999
- [Sawik, 1996] Sawik T.: Planowanie i sterowanie produkcji w elastycznych systemach montazowych, WNT, Warszawa, 1996
- [Setlak, 2004] G.Setlak. Harmonogramowanie produkcji na wydziale montażu za pomocą algorytmów ewolucyjnych // Ogólnopolski kwartalnik naukowo-techniczny „Technologia i automatyzacja montażu”, 2004, № 2, P..3-9.
- [Instruction, 1998] The instruction of disassembly and repairs, the catalogue of teams and the spare parts for vacuum cleaner type 900.0 i 930.0 ZZSD „PREDOM-ZELMER”, Rzeszow, 1998.
- [Uckun, 1993] Uckun S., Bagchi S., Mijabe Y.: Managing genetic search in Job-shop scheduling //IEEE Expert-Intelligent Systems & Their Applications, Vol.8, 1993, pp.15-24.

---

**Authors' Information**

---

**Galina Setlak** – D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science, Str. W. Pola 2 Rzeszow 35-959, and The Bronislaw Markiewicz State School Of Higher Vocational Education in Jaroslaw, Czarnieckiego Street 16, Poland, Phone: (48-17)- 86-51-433, e-mail: [gsetlak@prz.edu.pl](mailto:gsetlak@prz.edu.pl)

## THE CASCADE NEO-FUZZY ARCHITECTURE AND ITS ONLINE LEARNING ALGORITHM

Yevgeniy Bodyanskiy, Yevgen Viktorov

**Abstract:** *in the paper learning algorithm for adjusting weight coefficients of the Cascade Neo-Fuzzy Neural Network (CNFNN) in sequential mode is introduced. Concerned architecture has the similar structure with the Cascade-Correlation Learning Architecture proposed by S.E. Fahlman and C. Lebiere, but differs from it in type of artificial neurons. CNFNN consists of neo-fuzzy neurons, which can be adjusted using high-speed linear learning procedures. Proposed CNFNN is characterized by high learning rate, low size of learning sample and its operations can be described by fuzzy linguistic “if-then” rules providing “transparency” of received results, as compared with conventional neural networks. Using of online learning algorithm allows to process input data sequentially in real time mode.*

**Keywords:** *artificial neural networks, constructive approach, fuzzy inference, hybrid systems, neo-fuzzy neuron, real-time processing, online learning.*

**ACM Classification Keywords:** *I.2.6 Learning – Connectionism and neural nets*

**Conference:** *The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009*

---

### Introduction

---

Nowadays artificial neural networks (ANNs) are widely applied for solving a large class of problems related with the processing of information given as time-series or numerical “object-property” tables generated by non-stationary, chaotic or stochastic systems. However in real conditions data processing often must be performed simultaneously with the plant functioning and therefore weight adaptation must be executed in a sequential mode as well. So called “optimization-based networks” such as Multilayer Perceptron, Radial Basis Functions Network (RBFN), Normalized Radial Basis Functions Network (NRBFN) in most cases cannot be effective during solving mentioned above problems because of their low convergence rate, curse of dimensionality, and impossibility to learn in on-line mode.

In the papers [Bodyanskiy, 2008a, Bodyanskiy, 2008b, Bodyanskiy, 2008c] we have introduced various modifications of so called cascade artificial neural networks [Fahlman, 1990; Schalkoff, 1997; Avedjan, 1999], which have variable growing architecture and differs by the type of nodes – artificial neurons. It was shown that using of neo-fuzzy neurons [Yamakawa, 1992; Uchino, 1997; Miki, 1999] as elementary structural components of the cascade networks gives such valuable advantages as high learning rate, low size of the learning sample, and possibility to describe overall artificial neural network functioning process by the fuzzy linguistic “if-then” rules, what provides transparency of received results and therefore increases the range of applications for this architecture. It should be noticed that listed advantages are common for entire class of hybrid neo-fuzzy systems [Jang, 1997].

But as it was stated above possibility to adjust synaptic weight coefficients of the network is quite attractive and even necessary in some cases technique. So at this paper an attempt of synthesis of such procedure which possesses both smooth and filtering properties is taken.

---

### The Neo-Fuzzy Neuron

---

Neo-fuzzy neuron is a nonlinear multi-input single-output system shown in Fig.1.

It realizes the following mapping:

$$\hat{y} = \sum_{i=1}^n f_i(x_i) \tag{1}$$

where  $x_i$  is the  $i$ -th input ( $i = 1, 2, \dots, n$ ),  $\hat{y}$  is a system output. Structural blocks of neo-fuzzy neuron are nonlinear synapses NS<sub>*i*</sub> which perform transformation of  $i$ -th input signal in the form

$$f_i(x_i) = \sum_{j=1}^h w_{ji} \mu_{ji}(x_i).$$

Each nonlinear synapse realizes the fuzzy inference

IF  $x_i$  IS  $x_{ji}$  THEN THE OUTPUT IS  $w_{ji}$

where  $x_{ji}$  is a fuzzy set which membership function is  $\mu_{ji}$ ,  $w_{ji}$  is a singleton (synaptic weight) in consequent [Uchino, 1997]. As it can be readily seen nonlinear synapse in fact realizes Takagi-Sugeno fuzzy inference of zero order.

Conventionally the membership functions  $\mu_{ji}(x_i)$  in the antecedent are complementary triangular functions. For preliminary normalized input variables  $x_i$  (usually  $0 \leq x_i \leq 1$ ), they can be expressed in the form:

$$\mu_{ji}(x_i) = \begin{cases} \frac{x_i - c_{j-1,i}}{c_{ji} - c_{j-1,i}}, & x \in [c_{j-1,i}, c_{ji}], \\ \frac{c_{j+1,i} - x_i}{c_{j+1,i} - c_{ji}}, & x \in [c_{ji}, c_{j+1,i}], \\ 0 & \text{otherwise} \end{cases}$$

where  $c_{ji}$  are arbitrarily selected centers of corresponding membership functions. Usually they are equally distributed on interval [0, 1]. This contributes to simplify the fuzzy inference process. That is, an input signal  $x_i$  activates only two neighboring membership functions simultaneously and the sum of the grades of these two membership functions equals to unity (Ruspini partitioning), i.e.

$$\mu_{ji}(x_i) + \mu_{j+1,i}(x_i) = 1.$$

Thus, the fuzzy inference result produced by the Center-of-Gravity defuzzification method can be given in the very simple form

$$f_i(x_i) = w_{ji} \mu_{ji}(x_i) + w_{j+1,i} \mu_{j+1,i}(x_i).$$

By summing up  $f_i(x_i)$ , the output  $\hat{y}$  of Eq. (1) is produced.

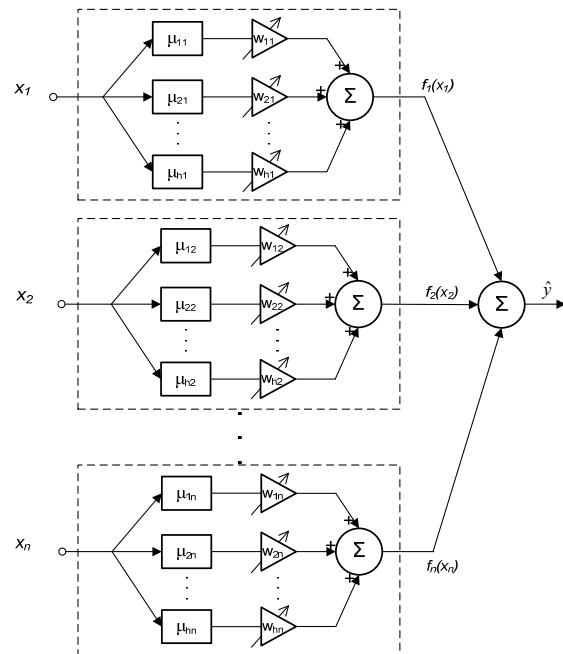


Figure 1. The Neo-Fuzzy Neuron

### The Cascade Neo-Fuzzy Neural Network Architecture

The Cascade Neo-Fuzzy Neural Network architecture shown in Fig.2 and mapping which it realizes has the following form:

- neo-fuzzy neuron of the first cascade

$$\hat{y}^{[1]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[1]} \mu_{ji}(x_i),$$

- neo-fuzzy neuron of the second cascade

$$\hat{y}^{[2]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[2]} \mu_{ji}(x_i) + \sum_{j=1}^h w_{j,n+1}^{[2]} \mu_{j,n+1}(\hat{y}^{[1]}),$$

- neo-fuzzy neuron of the third cascade

$$\hat{y}^{[3]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[3]} \mu_{ji}(x_i) + \sum_{j=1}^h w_{j,n+1}^{[3]} \mu_{j,n+1}(\hat{y}^{[1]}) + \sum_{j=1}^h w_{j,n+2}^{[3]} \mu_{j,n+2}(\hat{y}^{[2]}),$$

- neo-fuzzy neuron of the  $m$ -th cascade

$$\hat{y}^{[m]} = \sum_{i=1}^n \sum_{j=1}^h w_{ji}^{[m]} \mu_{ji}(x_i) + \sum_{l=n+1}^{n+m-1} \sum_{j=1}^h w_{jl}^{[m]} \mu_{jl}(\hat{y}^{[l-n]}). \quad (2)$$

Thus cascade neo-fuzzy neural network contains  $h \left( n + \sum_{l=1}^{m-1} l \right)$  adjustable parameters and it is important that all of them are linearly included in the definition (2).

Let us define  $h(n+m-1) \times 1$  membership functions vector of  $m$ -th neo-fuzzy neuron  $\mu^{[m]} = (\mu_{11}(x_1), \dots, \mu_{h1}(x_1), \mu_{12}(x_2), \dots, \mu_{h2}(x_2), \dots, \mu_{ji}(x_i), \dots, \mu_{hn}(x_n), \mu_{1,n+1}(\hat{y}^{[1]}), \dots, \mu_{h,n+1}(\hat{y}^{[1]}), \dots, \mu_{h,n+m-1}(\hat{y}^{[m-1]}))^T$  and corresponding vector of synaptic weights  $w^{[m]} = (w_{11}^{[m]}, w_{21}^{[m]}, \dots, w_{h1}^{[m]}, w_{12}^{[m]}, \dots, w_{h2}^{[m]}, \dots, w_{ji}^{[m]}, \dots, w_{hn}^{[m]}, w_{1,n+1}^{[m]}, \dots, w_{h,n+1}^{[m]}, \dots, w_{h,n+m-1}^{[m]})^T$  which has the same dimensionality. Then we can represent expression (2) in vector notation:

$$\hat{y}^{[m]} = w^{[m]T} \mu^{[m]}.$$

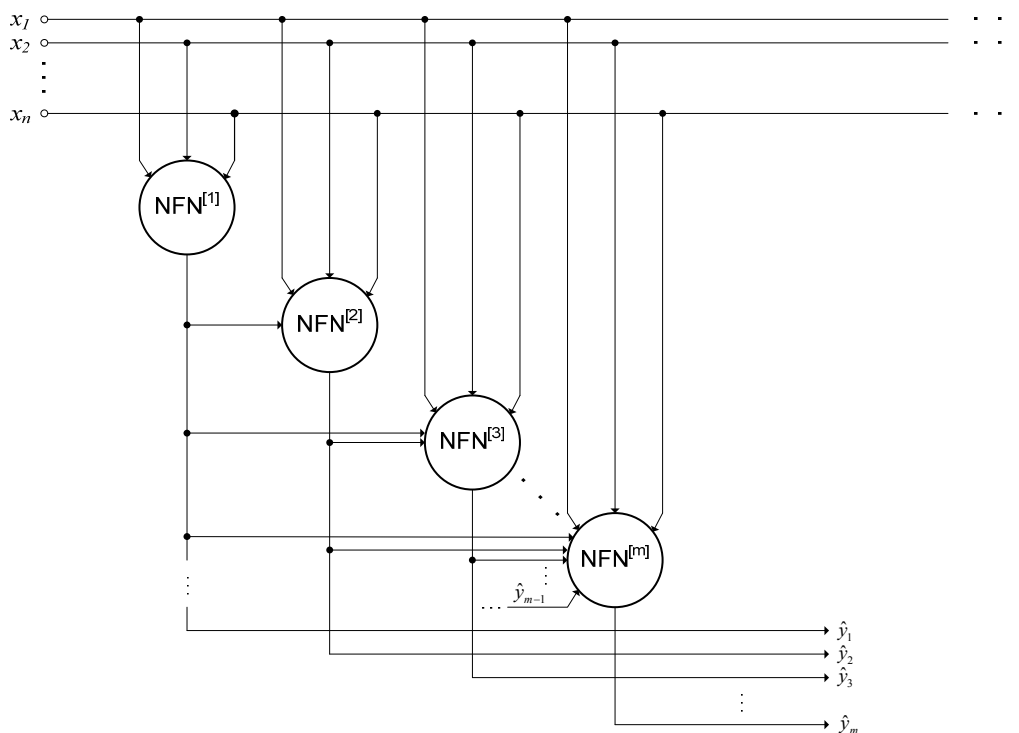


Figure 2. The Cascade Neo-Fuzzy Neural Network Architecture.

### The Cascade Neo-Fuzzy Neural Network Sequential Learning Algorithm

Learning algorithm for the cascade neo-fuzzy architecture in general form can be found in [Bodyanskiy, 2008c]. It is said there that network's growing process (increasing quantity of cascades) continues until we obtain required precision of the solved problem's solution, and for adjusting weight coefficients of the last  $n$ -th cascade following expressions are used:

$$w^{[m]}(N) = \left( \sum_{k=1}^N \mu^{[m]}(k) \mu^{[m]T}(k) \right)^+ \sum_{k=1}^N \mu^{[m]}(k) y(k) = P^{[m]}(N) \sum_{k=1}^N \mu^{[m]}(k) y(k) \quad (3)$$

in a batch mode or

$$\begin{cases} w^{[m]}(k+1) = w^{[m]}(k) + \frac{P^{[m]}(k)(y(k+1) - w^{[m]T}(k)\mu^{[m]}(k+1))}{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)} \mu^{[m]}(k+1), \\ P^{[m]}(k+1) = P^{[m]}(k) - \frac{P^{[m]}(k)\mu^{[m]}(k+1)\mu^{[m]T}(k+1)P^{[m]}(k)}{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)}, P^{[m]}(0) = \beta I \end{cases} \quad (4)$$

or

$$\begin{cases} w^{[m]}(k+1) = w^{[m]}(k) + (r^{[m]}(k+1))^{-1} (y(k+1) - w^{[m]T}(k)\mu^{[m]}(k+1)) \mu^{[m]}(k+1), \\ r^{[m]}(k+1) = \alpha r^{[m]}(k) + \|\mu^{[m]}(k+1)\|^2, 0 \leq \alpha \leq 1 \end{cases} \quad (5)$$

in a serial mode.

It should be noticed that in general case algorithms (3) and (4) are not coincident since in (3)

$$P^{[m]}(k) = \left( \sum_{k=1}^N \mu^{[m]}(k) \mu^{[m]T}(k) \right)^+$$

and in (4)

$$P^{[m]}(k) = \left( \sum_{k=1}^N \mu^{[m]}(k) \mu^{[m]T}(k) \right)^{-1}.$$

If matrix  $\sum_{k=1}^N \mu^{[m]}(k) \mu^{[m]T}(k)$  is singular or ill-conditioned, algorithm (4) becomes nonoperatable. And in case we use adjusting additions  $P^{[m]}(0) = \beta I$ , synaptic weight coefficients estimations can be significantly inaccurate and biased. Using Greville's theorem in pseudoinversion procedure allows to write algorithm (4) in more general form [Albert, 1972]:

$$w^{[m]}(k+1) = w^{[m]}(k) + \gamma^{[m]}(k+1) (y(k+1) - w^{[m]T}(k)\mu^{[m]}(k+1)) \mu^{[m]}(k+1), \quad (6)$$

$$\Gamma^{[m]}(k+1) = \begin{cases} \frac{Q^{[m]}(k)}{\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)}, & \text{if } \mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1) \geq \varepsilon(k+1), \\ \frac{P^{[m]}(k)}{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)}, & \text{if not,} \end{cases} \quad (7)$$

$$Q^{[m]}(k+1) = \begin{cases} Q^{[m]}(k) - \frac{Q^{[m]}(k)\mu^{[m]}(k+1)\mu^{[m]T}(k+1)Q^{[m]}(k)}{\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)}, & \text{if } \mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1) \geq \varepsilon(k+1), \\ Q^{[m]}(k), & \text{if not,} \end{cases} \quad (8)$$

$$P^{[m]}(k+1) = \begin{cases} P^{[m]}(k) - \frac{\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)\left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)^T + \left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)^T}{\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)} + \\ + \frac{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)\left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)^T}{\left(\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)\right)^2}, \\ \text{if } \mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1) \geq \varepsilon(k+1), \\ P^{[m]}(k) - \frac{P^{[m]}(k)\mu^{[m]}(k+1)\mu^{[m]T}(k+1)P^{[m]}(k)}{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)}, \text{ if not} \end{cases} \quad (9)$$

where  $\varepsilon(k+1)$  - nonnegative threshold which defines degree of vectors  $\mu^{[m]}(k+1)$  multi-collinearity and designates appropriate processing method.

Advantages of procedure (6)-(9) are numerical stability and possibility to perform network learning when number of observations  $N$  is lesser then number of parameters which should be estimated  $h(n+m-1)$ .

In case we have deal with nonstationary data, when parameters of required solution unpredictably vary with time, algorithms with exponential reducing of information value can be used, for example gradient procedure (5). If tracking speed of gradient algorithm isn't sufficient, second order procedures can be utilized as well, for example exponentially weighted recurrent least squares method in form [Ljung, 1987]:

$$\begin{cases} w^{[m]}(k+1) = w^{[m]}(k) + \frac{P^{[m]}(k)\left(y(k+1) - w^{[m]T}(k)\mu^{[m]}(k+1)\right)}{\alpha + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)}\mu^{[m]}(k+1), \\ P^{[m]}(k+1) = \frac{1}{\alpha} \left( P^{[m]}(k) - \frac{P^{[m]}(k)\mu^{[m]}(k+1)\mu^{[m]T}(k+1)P^{[m]}(k)}{\alpha + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)} \right), \quad 0 < \alpha \leq 1. \end{cases} \quad (10)$$

It should be noticed that usage of algorithm (10) can lead to so called covariance matrix  $P^{[m]}(k+1)$  "parameters blow-up", i.e. exponential growth of its elements. This can be avoided using valid forgetting parameter  $\alpha$ , which usually selected in short range  $0.95 \leq \alpha \leq 0.99$ . Decreasing  $\alpha$  value results in rapid matrix  $P^{[m-1]}(k+1) = \sum_{l=1}^{k+1} \alpha^{k+1-l} \mu^{[m]}(l)\mu^{[m]T}(l)$  degeneration and therefore "parameters blow-up". Usage of pseudoinverse procedure based on Greville's theorem in algorithm (10) gives learning procedure [Bodyanskiy, 1985, Bodyanskiy, 1996, Bodyanskiy, 1998]:

$$w^{[m]}(k+1) = w^{[m]}(k) + \gamma^{[m]}(k+1)\left(y(k+1) - w^{[m]}(k)\mu^{[m]}(k+1)\right)\mu^{[m]}(k+1), \quad (11)$$

$$\Gamma^{[m]}(k+1) = \begin{cases} \frac{Q^{[m]}(k)}{\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)}, \text{ if } \mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1) \geq \varepsilon(k+1), \\ \frac{P^{[m]}(k)}{\alpha + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)}, \text{ if not,} \end{cases} \quad (12)$$

$$P^{[m]}(k+1) = \begin{cases} \frac{1}{\alpha} \left( P^{[m]}(k) - \frac{\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)\left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)^T + \left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)^T}{\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)} + \right. \\ \left. + \frac{1 + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)\left(Q^{[m]}(k)\mu^{[m]}(k+1)\right)\left(P^{[m]}(k)\mu^{[m]}(k+1)\right)^T}{\left(\mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1)\right)^2} \right), \\ \text{if } \mu^{[m]T}(k+1)Q^{[m]}(k)\mu^{[m]}(k+1) \geq \varepsilon(k+1), \\ \frac{1}{\alpha} \left( P^{[m]}(k) - \frac{P^{[m]}(k)\mu^{[m]}(k+1)\mu^{[m]T}(k+1)P^{[m]}(k)}{\alpha + \mu^{[m]T}(k+1)P^{[m]}(k)\mu^{[m]}(k+1)} \right), \text{ if not} \end{cases} \quad (13)$$

(here  $Q^m(k)$ ,  $Q^{[m]}(k+1)$  are defined by expression (8)). Proposed procedure is stable in any value of forgetting parameter. It can be seen that procedure given by equations (8), (11)-(13) is a generalization of algorithms (4), (6)-(9), and (10).

## Simulation Results

In order to confirm the performance of the proposed architecture the prediction of time-series is examined. We applied the proposed algorithm which allows to perform Cascade Neo-Fuzzy Neural Network learning in sequential mode for the forecasting of a chaotic process defined by the Mackey-Glass equation [Mackey, 1977]:

$$y'(t) = \frac{0,2t(t-\tau)}{1+y^{10}(t-\tau)} - 0,1y(t). \quad (14)$$

The signal defined by (14) was quantized with step 0.1. We took a fragment containing 700 points. The goal was to predict time-series value on the next step  $k+1$  using its values on steps  $k-3$ ,  $k-2$ ,  $k-1$ , and  $k$ . To bring into obtained set of signal values additional nonstationarity to several intervals different positive or negative numbers were added. First 500 points were used to adjust weight coefficients of the cascade architecture in sequential mode. It means that during learning procedure artificial neural network already performed time-series prediction beginning from the first element which was fed to it. Remaining 200 points were processed by cascade architecture without adjusting its weight coefficients.

For simulation modeling cascade network which consists of three cascades was synthesized. Each cascade contained single neo-fuzzy neuron with three activation functions. Overall quantity of parameters which should be determined was 45. We used  $\alpha = 0.985$  during weight adaptation procedure in algorithm (8), (11)-(13).

For estimation of received results normalized mean square error (NRMSE) as well as mean square (MSE) error was used. Obtained results of Mackey-Glass time-series prediction are shown in Fig. 3.

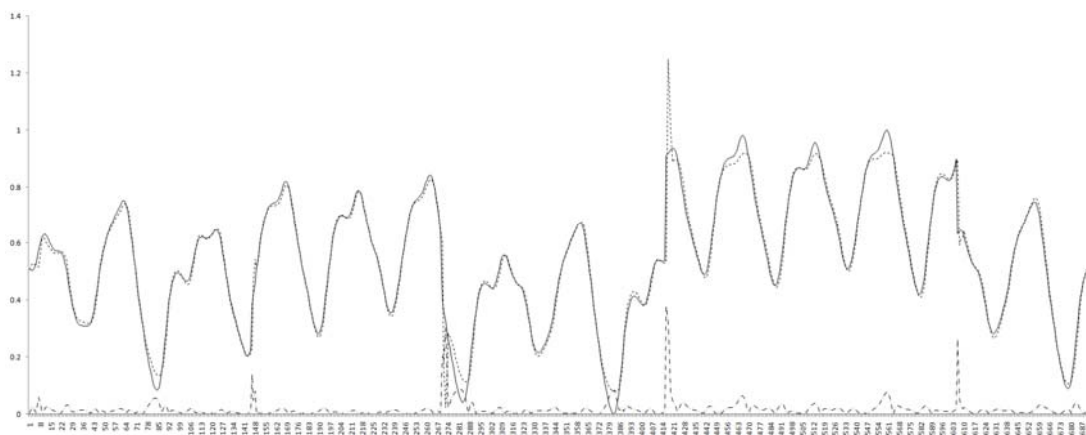


Figure 3. Mackey-Glass time-series prediction:  
original signal – solid line; network output – dashed line; prediction error – chain line.

Calculated on described dataset errors were the following:  $MSE = 0.008$ ,  $NRMSE = 0.3$ . As it can be readily seen from the figure signal changed its y-coordinate center four times. At each such case temporary prediction error burst occurred. Its magnitude depended on the nonstationarity power, and it's obviously that the more drastic changes of predicted signal take place the greater error burst is happen. But after several examples are fed to network input and synaptic weight coefficients became adjusted according to changed time-series, further prediction process flows quite well.

In whole, proposed algorithm for Cascade Neo-Fuzzy Neural Network gives very close approximation and prediction quality of sufficiently nonstationary processes in online mode.

---

## Conclusion

---

The new algorithm for Cascade Neo-Fuzzy Neural Network which allows to perform synaptic weight adaptation in sequential mode is proposed. It gives opportunity to start the prediction process from the first element which was fed to network's input irrespectively from the quantity of parameters which should be determined. Theoretical justification and experiment results confirm the efficiency of developed approach.

---

## Acknowledgements

---

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

## Bibliography

---

- [Bodyanskiy, 2008a] Bodyanskiy Ye., Dolotov A., Pliss I., Viktorov Ye. The cascade orthogonal neural network. Advanced Research in Artificial Intelligence. Bulgaria, Sofia: Institute of Informational Theories and Applications FOI ITHEA, 2. 2008. P. 13-20.
- [Bodyanskiy, 2008b] Bodyanskiy Ye., Viktorov Ye. Time-series prediction using cascade orthogonal neural network. Radioelektronika. Informatika. Upravlinnja. №1. 2008. P. 92-97.
- [Bodyanskiy, 2008c] Bodyanskiy Ye., Viktorov Ye., Pliss I. The cascade neo-fuzzy neural network and its learning algorithm. Visnyk Uzhgorods'kogho Nacional'nogho universytetu. Serija "Matematyka i informatyka", Issue 17. 2008. P. 48-58.
- [Fahlman, 1990] Fahlman S.E., Lebiere C. The cascade-correlation learning architecture. In: Advances in Neural Information Processing Systems. Ed. D.S. Touretzky. San Mateo, CA: Morgan Kaufman, 1990. P. 524-532.
- [Schalkoff, 1997] Schalkoff R.J. Artificial Neural Networks. N.Y.: The McGraw-Hill Comp., 1997.
- [Avedjan, 1999] Avedjan E.D., Barkan G.V., Levin I.K. Cascade neural networks. Avtomatika i Telemekhanika, 3. 1999. P. 38-55.
- [Yamakawa, 1992] Yamakawa T., Uchino E., Miki T., Kusanagi H. A neo fuzzy neuron and its applications to system identification and prediction of the system behavior. Proc. of 2-nd Int.Conf. on Fuzzy Logic and Neural Networks "IIZUKA-92". Iizuka, Japan, 1992. P. 477-483.
- [Uchino, 1997] Uchino E., Yamakawa T. Soft computing based signal prediction, restoration and filtering. Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms. Ed. Da Ruan. Boston: Kluwer Academic Publisher, 1997. P. 331-349.
- [Miki, 1999] Miki T., Yamakawa T. Analog implementation of neo-fuzzy neuron and its on-board learning. Computational Intelligence and Applications. Ed. N. E. Mastorakis. Piraeus: WSES Press, 1999. P. 144-149.
- [Jang, 1997] Jang Jr. S. R., Sun C. T., Mizutani E. Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence. N.J.: Prentice Hall, 1997. 614p.
- [Albert, 1972] Albert A. Regression and the Moore-Penrose Pseudoinverse. New York & London, Academic Press, 1972.
- [Ljung, 1987] Ljung L. System Identification: Theory for the User. N.J., Englewood Cliffs, Prentice-Hall. 1987.
- [Bodyanskiy, 1985] Bodyanskiy Ye. V., Buryak Y. O., Sodin M. L. Adaptive algorithms of identification with finite memory. Kharkiv, ruk. dep. v UkrNIINTI 05.02.1985, №218UK, 85 Dep. 1984. 47p.
- [Bodyanskiy, 1996] Bodyanskiy Ye. V., Kotlyarevskiy S. V., Achkasov A. Ye., Voronovskiy G. K. Low Order Adaptive Regulators. Kharkiv, KhGAGKh. 1996.
- [Bodyanskiy, 1998] Bodyanskiy Ye. V., Rudenko O. G. Adaptive Models in Technical Plants Control Systems. Kyiv, UMK VO. 1998.
- [Mackey, 1977] Mackey M. C., Glass L. Oscillation and chaos in physiological control systems. Science, №197. 1977. P. 238-289.

---

## Authors' Information

---

**Yevgeniy Bodyanskiy** – Professor, Dr.-Ing. habil., Professor of Artificial Intelligence Department and Scientific Head of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Lenina av. 14, Kharkiv, 61166, Ukraine, Tel +380577021890, e-mail: [bodya@kture.kharkov.ua](mailto:bodya@kture.kharkov.ua)

**Yevgen Viktorov** - Ph.D. Student, Kharkiv National University of Radio Electronics, Lenin Av., 14, Kharkiv, 61166, Ukraine, Tel +380681613429, e-mail: [yevgen.viktorov@gmail.com](mailto:yevgen.viktorov@gmail.com)



---

## ANALYSIS OF P-SYSTEMS UNDER A MULTIAGENT SYSTEMS PERSPECTIVE

**Alberto Arteta, Angel Goñi, Juan Castellanos**

**Abstract:** Membrane computing is a recent area that belongs to natural computing. This field works on computational models based on nature's behavior to process the information. Recently, numerous models have been developed and implemented with this purpose. P-systems are the structures which have been defined, developed and implemented to simulate the behavior and the evolution of membrane systems which we find in nature. What we show in this paper is an application capable to simulate the P-systems based on a multiagent systems (MAS) technology. The main goal we want to achieve is to take advantage of the inner qualities of the multiagent systems. This way we can analyse the proper functioning of any given p-system. When we observe a P-system from a different perspective, we can be assured that it is a particular case of the multiagent systems. This opens a new possibility, in the future, to always evaluate the P-systems in terms of the multiagent systems technology.

**Keywords:** P-systems mapping, multiagent systems.

**Conference:** The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009

---

### Introduction

---

Natural computing is a new field within computer science which develops new computational models. These computational models can be divided into three major areas:

- Neuronal networks.
- Genetic Algorithms
- Biomolecular computation.

Membrane computing is included in biomolecular computation. Within the field of membrane computing a new logical computational device appears: The P-system. These P-systems are able to simulate the behavior of the membranes on living cells. This behavior refers to the way membranes process information. (Absorbing nutrients, chemical reactions, dissolving, etc)

Membrane computing formally represents, through the use of P-systems, the processes that take place inside of the living cells. In terms of software systems, it is the process within a complex and distributed software. In parallel computational models, p-systems might be as important as the Turing machine is in sequential computational models.

In this paper, we design a multiagent system that it is capable to simulate the behavior of any given P-system. After this we will study the advantages of applying this system. By simulating the P-systems with a multiagent system, we can analyze the P-system from a different perspective. Therefore establishing a procedure that will focus on the correction of any given P-system based on the multiagent system paradigm.

In order to do this we will take the following steps:

- Introduction to P-systems theory;
- Components of a particular multiagent system;
- Link between P-systems and multiagent systems;
- Example and code;
- Conclusions and further work.

## Introduction to P-systems theory

In this section we will study into detail all of the theories related to the paradigm of the P-systems. A P-system is a computational model inspired by the way the living cells interact with each other through their membranes. The elements of the membranes are called objects. A region within a membrane can contain objects or other membranes. A p-system has an external membrane (also called skin membrane) and it also contains a hierarchical relation defined by the composition of the membranes. A multiset of objects is defined within a region (enclosed by a membrane). These multisets of objects show the number of objects existing within a region. Any object 'x' will be associated to a multiplicity which tells the number of times that 'x' is repeated in a region.

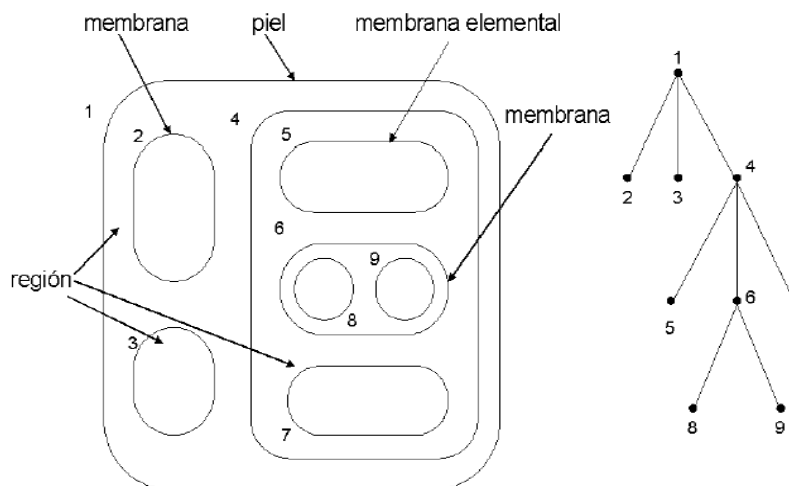


Fig. 1. The membrane's structure (left) represented in tree shape (right)

According to Păun 's definition, a transition P System of degree  $n$ ,  $n > 1$  is a construct: [Păun 1998]

$$\Pi = (V, \mu, \omega_1, \dots, \omega_n, (R_1, \rho_1), \dots, (R_n, \rho_n), i_0)$$

where:

1.  $V$  is an alphabet; its elements are called objects;
2.  $\mu$  is a membrane structure of degree  $n$ , with the membranes and the regions labeled in a one-to-one manner with elements in a given set ; in this section we always use the labels  $1, 2, \dots, n$ ;
3.  $\omega_i$   $1 \leq i \leq n$  , are strings from  $V^*$  representing multisets over  $V$  associated with the regions  $1, 2, \dots, n$  of  $\mu$
4.  $R_i$   $1 \leq i \leq n$  , are finite set of evolution rules over  $V$  associated with the regions  $1, 2, \dots, n$  of  $\mu$ ;  $\rho_i$  is a partial order over  $R_i$   $1 \leq i \leq n$  , specifying a priority relation among rules of  $R_i$  . An evolution rule is a pair  $(u, v)$  which we will usually write in the form  $u \rightarrow v$  where  $u$  is a string over  $V$  and  $v = v'$  or  $v = v' \delta$  where  $v'$  is a string over  $(V \times \{here, out\}) \cup (V \times \{in_j \mid 1 \leq j \leq n\})$ , and  $\delta$  is a special symbol not in. The length of  $u$  is called the radius of the rule  $u \rightarrow v$
5.  $i_0$  is a number between 1 and  $n$  which specifies the output membrane of  $\Pi$

Let  $U$  be a finite and not an empty set of objects and  $N$  the set of natural numbers. A *multiset of objects* is defined as a mapping:

$$M : V \rightarrow N$$

$$a_i \rightarrow u_i$$

Where  $a_i$  is an object and  $u_i$  its multiplicity.

As it is well known, there are several representations for multisets of objects.

$$M = \{(a_1, u_1), (a_2, u_2), (a_3, u_3), \dots\} = a_1^{u_1} \cdot a_2^{u_2} \cdot a_n^{u_n} \dots$$

Evolution rule with objects in  $U$  and targets in  $T$  is defined by  $r = (m, c, \delta)$

where  $m \in M(V)$ ,  $c \in M(V \times T)$  and  $\delta \in \{\text{to dissolve, not to dissolve}\}$

From now on 'c' will be referred to as the consequent of the evolution rule 'r'

The set of evolution rules with objects in  $V$  and targets in  $T$  is represented by  $R(U, T)$ .

We represent a rule as:

$x \rightarrow y$  or  $x \rightarrow y\delta$  where  $x$  is a multiset of objects in  $M((V) \times \text{Tar})$  where  $\text{Tar} = \{\text{here, in, out}\}$  and  $y$  is the consequent of the rule. When  $\delta$  is equal to "dissolve", then the membrane will be dissolved. This means that objects from a region will be placed within the region which contains the dissolved region. Also, the set of evolution rules included on the dissolved region will disappear.

P-systems evolve, which makes it change upon time; therefore it is a dynamic system. Every time that there is a change on the p-system we will say that the P-system is in a new transition. The step from one transition to another one will be referred to as an evolutionary step, and the set of all evolutionary steps will be named computation. Processes within the p-system will be acting in a massively parallel and non-deterministic manner. (Similar to the way the living cells process and combine information).

We will say that the computation has been successful if:

- The halt status is reached.
- No more evolution rules can be applied.
- Skin membrane still exists after the computation finishes.

---

## Multiagent System Technology

---

The multi agent system we are going to study is a system made of several entities of computers. These systems can achieve goals unreachable for single agent systems (one agent system). The main characteristics of multiagent systems are:

- Proactivity;
- Autonomy.

Formally an agent is a real or virtual entity that:

1. Is able to act within a given environment.
2. Is able to communicate with other agents.
3. It has its own resources.
4. Is able to retrieve information and to (at least partially) know the environment.
5. It can reproduce.

According to the definition of agent, a multiagent system can be defined as a system of computers which contains the following elements:

1. An environment  $E$ , which is a space.
2. A set of objects  $O \in E$ .
3. A set of agents  $A \in O$ .
4. A set of relations  $R$  between objects and agents.
5. A set of operations that allows the agents interact with the objects.

When studying the P-system under a multiagent system technology we will consider  $A=O$ . All of our agents will relate with each other and will communicate with each other.

In the next sections we will see that p-systems can be studied under multiagent systems technology by performing some transformations. The goal of this paper is to provide more properties to the p-system than the ones they have nowadays.

---

### Link from P-systems to Multiagent Systems

---

Once we have seen the structures of p-systems and multiagent systems separately, we just need to establish a link between these two universes. Once we build this link from one universe to the other, we will be able to see p-systems under a different perspective and therefore treat them differently.

Due to the fact that a p-system and a multiagent system have several components, it makes sense trying to establish a relation between the two of them. Thus, the first step is changing the perspective of the membranes. The way to do this is:

In a p-system, given a set of membranes  $M = \{m_i \mid i \in \mathbb{N}, 1 \leq i \leq n\}$  where  $m_i$  is a membrane, we establish a function  $f_{agent}$  which is defined as:

$$f_{agent} : M \rightarrow A$$

$$f_{agent}(m_i) = a_i \quad \forall i \in \mathbb{N} \quad i \leq n, \quad n \text{ number of membranes}$$

Under this new perspective, every agent  $a_i \in A$  has a set of resources which includes:

- Multiset of objects within the region enclosed by the membrane  $m_i \in M$
- Set of evolution rules within the region enclosed by the membrane of  $m_i \in M$

Besides, each membrane  $m_i \in M$  which contains another membrane  $m_j \in M$  can be mapped as father-son relationship within a multiagent system.

In order to set up the synchronism in our system, we need to have an agent called  $a_{sync} \in A$ . This agent will make sure that the membranes are synchronized.

The p-system will consider three major stages:

1. Static structure of the p-system
2. Dynamic behavior of the p-system
3. Synchronism between membranes.

Now is when we can describe the mapping:

Given a set of membranes  $M = \{m_i \mid i \in \mathbb{N}, 1 \leq i \leq n\}$  and  $W = \{\omega_i \mid i \in \mathbb{N}, 1 \leq i \leq n\}$  the set of multisets of a p-system where  $\omega_i \in W$  is the multiset of objects representing atoms, molecules included in the membrane  $m_i \in M$ . Let  $R_i$  be a finite and non empty set of evolution rules belonging to the membrane  $m_i \in M$  and let  $u$  be the structure of membranes that define a father-son relation between membranes.

We can see a membrane  $m_i \in M$  as  $m_i = (\omega_i, R_i, children)$  where  $\omega_i$  is the multiset of objects and  $R_i$  the set of evolution rules. *Children* are the membranes which are contained in  $m_i \in M$  and also *children* are a subset of the set of membranes  $M$ .

The mapping between p-systems and multiagent systems can be built from a function and a proper operator:

$$f : M \times \{\omega_1, \omega_2, \dots, \omega_n\} \times \{R_1, R_2, \dots, R_n\} \rightarrow MAS$$

Where MAS is the multiagent system associated to the p-system. This multiagent system contains the agents  $a_i$ , resources  $R_i$  and a set of operations  $Op_i$

This function  $f$  can be defined in terms of 3 other sub-functions. These three sub-functions are:

- $$f_{name}: \mathbb{N} \rightarrow Agentname$$
1.  $f_{name}(i) = a_i$
  2.  $f_{resource}: \{\omega_1, \omega_2, \dots, \omega_n\} \times \{R_1, R_2, \dots, R_n\} \rightarrow Resources$
  3.  $f_{operation}: dynamic\ transition \rightarrow Agent\ behavior$

Every agent  $a_i$  will be associated to a set of resources called  $Res_i$  and set of operations called  $a_i$ . When referring to  $a_i$  it will mean  $a_i$  with its resources and operations.

Formally speaking, the multiagent system associated to a p-system with  $n$  membranes will have the set of resources as the union of all the objects, and the union of all the set of evolution rules included in every

membrane i.e.  $Resources = \left[ \bigcup_{i=1}^n R_i \right] \cup V$  where  $R_i$  is the set of evolution rules included in the membrane  $i$ .

By using this information we can provide more details of  $f_{resource}$

$f_{resource}$  will convert a multiset of objects  $\omega$  in a string like this:

The objects are represented as  $(x_i, n_i)$ , where  $x_i$  is the object and  $n_i$  its multiplicity.

Thus,  $f_{resource}$  will convert a multiset of objects:

$\omega = (x_1, n_1), (x_2, n_2), \dots, (x_n, n_m) \Rightarrow f_{resource}(\omega) = x_1 x_1 \dots x_1 x_2 x_2 \dots x_2 \dots$  Where every object  $x_i$  appears  $n_i$  times.

And given an evolution rule:

$$r = (x_1, n_1), (x_2, n_2), \dots, (x_n, n_m) \rightarrow (x_1, m_1), (x_2, m_2), \dots, (x_n, m_m)$$

$$\Rightarrow f_{resource}(r) = (x_1 \cdot x_1 x_2 \cdot x_2 \dots x_n \cdot x_n, x_1 \cdot x_1 x_2 \cdot x_2 \dots x_n \cdot x_n)$$

where  $n_i$  is the multiplicity of the object  $x_i$  in the antecedent and the multiplicity  $m_i$  in the consequent.  $f_{resource}(r)$ , has a two dimensional image. The first component of the image is a conversion of the antecedent into a string where every object  $x_i$  appears  $n_i$  times. The second component of the image is a conversion of the consequent into a string where every object  $y$   $x_i$  appears  $m_i$  times,  $n_i, m_i \in \mathbb{N}$ .

Once the mapping of the static parts has been developed, we are going to establish the necessary operations for the multiagent system to evolve in the same way that its associated p-system does. In order to achieve this we create an operator that returns the status of the transition of a p-system to a specific time. This synchronization must take place, and therefore it will have to be a part of the set of resources of the multiagent system. In order to study this synchronization, we create a new resource called *sync*. This resource is:

1. An integer (computing step)
2. A letter (Status)

The initial transition status is the integer 0.

$sync_i = \{\text{transition status of the agent } a_i\}$

The possible status in every computing step could be:

- A. Rules election,
- B. Objects consumption,.

**C. Communication between membranes**

We will also create an agent called  $a_{sync}$  which will be in charge of assuring the synchronization along with the agents (associated to the membranes).

For instance, let us have a set of transition statuses:

$$sync_1 = 3A \quad sync_2 = 2B, \quad sync_3 = 4C$$

Our agent  $a_{sync}$  will make sure this situation never happens. It will assure the following:

$$sync_i = sync_j \quad \forall i \neq j \quad i, j \in N$$

At last, the set of the union of all possible transition statuses will be part of the set of resources of the multiagent system.

Initially  $sync_i = 0 \quad \forall i \leq n \quad i \in N$ .

So far, we have seen that components of p-systems are mapped to multi agent system components (static part of a given p-system). Moreover, the evolution of a transition status with membrane systems can be also mapped as an evolution process within the multiagent systems. Because of this, we can state that modelling membrane systems can be seen as a multiagent system. This multiagent system has objects, agents, resources and operations.

Let us have three membranes  $m_1, m_2, m_3$  and  $m_1$  contains  $m_2$  which contains  $m_3$ . In our multiagent system this situation we would have three agents  $a_1, a_2, a_3$  where  $a_1$  is in the father-son relationship with  $a_2$  which is in the father-son relationship with  $a_3$ . It is represented as follows:

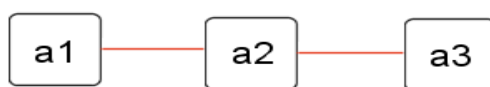


Fig. 2. Three membranes  $m_1$  contains  $m_2$  which contains  $m_3$  in the MAS.

Another way of representing the transformation of a set of membranes into MAS is a three of agents. In the three the root nodes are the fathers. The level below indicates the sons which can also be fathers of several agents. Given a generic set of membranes the mapping into MAS is as follows:

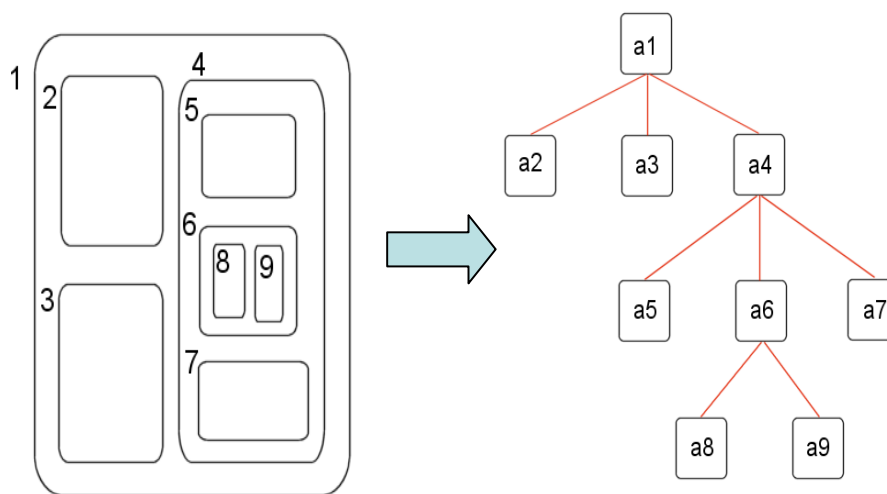


Fig. 3. Representation of a set of membranes into MAS.

in the way that a p-system evolves, it would be performed similarly in the MAS. In every evolution step the agent  $a_{sync}$  makes sure that the transition status of the membranes are the same. This assures the system evolves properly. In our MAS we will also have the following possible status:

1. Evolution rules selection
2. Evolution rules application
3. Communication between membranes

Example and code:

Once is clear the definitions and meanings of the P-systems and MAS and it is also clear the way we build the link between them we are going to show the transformation of a p-system into a MAS. So whenever an action occurs in a given p-system there will be a parallel action occurring in the MAS.

Let us suppose we have a p-system containing three membranes. This p-system is able to calculate the square of a given number. We are going to see 2 evolution steps (the initial one and the final one) in the p-systems can be transformed into MAS. To understand p-systems in detail, refer to [Arroyo, 2001].

a) Static components mapping

The P-system has:

- A set of membranes  $M = \{m_1, m_2, m_3\}$
- An Alphabet  $V = \{a, b, c, d, e, f\}$
- A set of multiset of objects  $M(V) = \{\omega_1 = \{\}, \omega_2 = \{\}, \omega_3 = \{af\}\}$  where  $\omega_i$  is the multiset of objects within the region delimited by the membrane  $m_i \quad \forall i \in \mathbb{N} \ i \leq 3$ .
- A Multiset of evolution rules

$$R(U, T) = \left\{ \begin{array}{l} R_1(U, T) = \{e \rightarrow e_{out}\}, R_2(U, T) = \{b \rightarrow d, d \rightarrow de, (ff \rightarrow f > f \rightarrow \delta)\}, \\ R_3(U, T) = \{a \rightarrow ab, a \rightarrow b\delta, f \rightarrow ff\} \end{array} \right\}$$

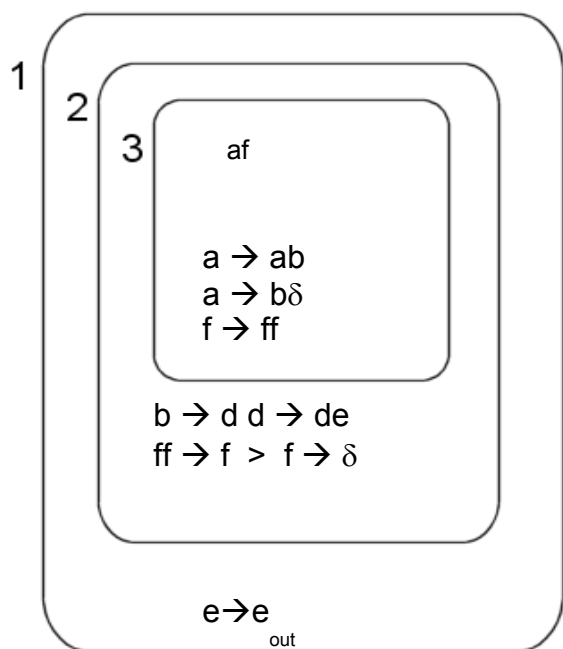


Fig 3 P-system which calculates a random square number.

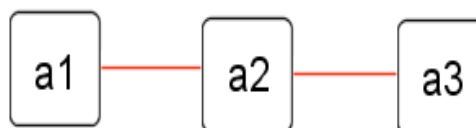


Fig 4. Associated Multisystem Agent.

The set of agents  $A = \{a_1, a_2, a_3, a_{sync}\}$

The resources used by the agent  $a_i$  are referred as  $Res_i$ , these are the multiset of objects and set of evolution rules.

$$\left\{ \bigcup_{i=1}^n Res_i \right\} = \left\{ \{[af], [(a, ab), (a, b\delta), (f, ff)]\}, \{[(b, d), (d, de), (ff, f), (f, \delta)]\}, \{[(e, e_{out})]\} \right\}$$

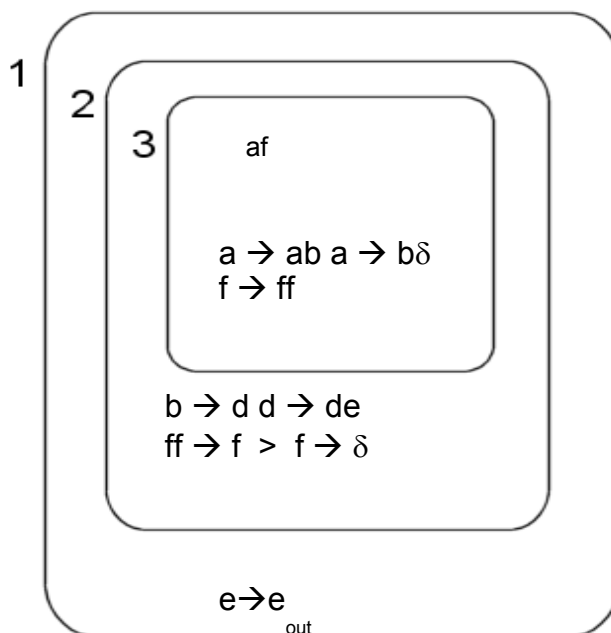
b) Dynamic behavior

In order to see how MAS evolves, we need to focus on:

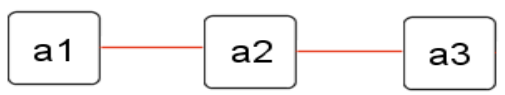
- Rules election
- Objects consumption
- Communication Stage

Initial Transition Status  $sync_i = 0 \quad \forall i \leq 3 \quad i \in \mathbb{N}$  This condition is checked by the agent  $a_{sync}$

In the transition status 1), the p-system evolves, In the Region 3 the rule number 1 and number three are applied.



Moreover, in our MAS:



our agent  $a_{sync}$  makes sure that  $sync_i = 1A \quad \forall i \leq 3 \quad i \in \mathbb{N}$ . In every region we are analysing the rules to be applied. Every agent selects from its resources the rules to be applied.

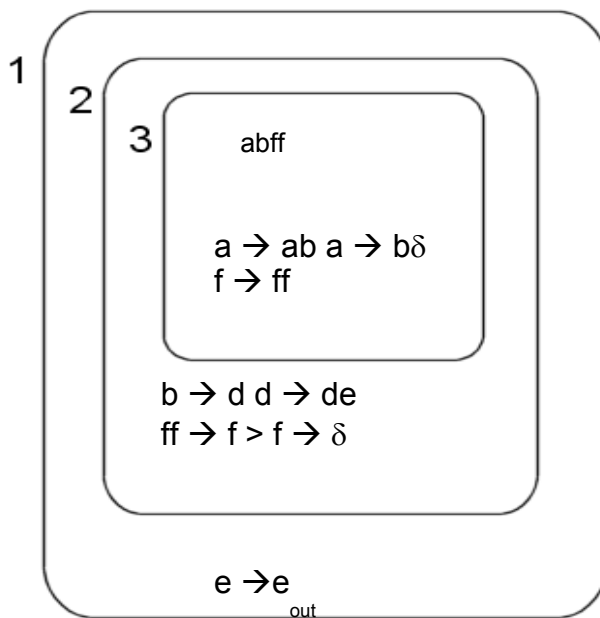
When this is done,  $a_{sync}$  assures we go to the next status.  $sync_i = 1B \quad \forall i \leq 3 \quad i \in \mathbb{N}$ . Here the agents execute the action “apply rules”.

Following our example, only agent  $a_3$  selects the rules  $r_1$  and  $r_3$  from its set of resources  $Res_3$ .

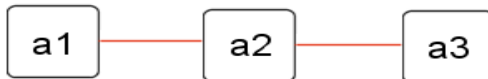


After applying rules  $a_{sync}$  assures all the agents are synchronized and go to the next status.  $sync_i = 1C \quad \forall i \leq 3 \quad i \in \mathbb{N}$ . Now is when the agents to communicate with each other and exchange objects if needed.

After the first computing step, we have:



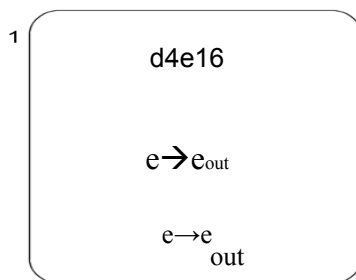
in our p-system, and our MAS is



where the resources used by the agents would be:

$$\left\{ \bigcup_{i=1}^n Res_i \right\} = \left\{ \{ [abff], [(a, ab), (a, b\delta), (f, ff)] \}, \{ [ [ (b, d), (d, de), (ff, f), (f, \delta) ] ] \}, \{ [ [ (e, e_{out}) ] ] \} \right\}$$

By repeating this process will reach a situation like the following:



Note. The number next to the object indicates the multiplicity of the object. which will return  $16=4^2$

In our MAS, we will end up with the following situation:



The rest of the agents have been eliminated when the action "dissolve" was performed.

$$\left\{ \bigcup_{i=1}^n Res_i \right\} = \{ \{ [d4e16], [e, e_{out}] \} \}$$

Code:

A possible code to implement the MAS associated to the p-system is:

*Rules R*

*Objects w*

*Dissolve = false*

*While NOT Dissolve*

*waitsync*

*//EVOLVE*

*R' = Rules\_election(w,R)*

*w' = Rules\_Application (w,R,P)*

*waitsync*

*//COMUNNICATION*

*w = COMMUNICATION(w')*

*dissolve = finish(w)*

*waitsync*: Synchronization between agents

*Rules\_election*: Selection of the rules

*Rules application*: Application of the rules

*communication (w)* Communication between agents and exchange of the objects w

*Finish*: The computation is finished

---

## Conclusion

---

In this paper, we have studied some topics of membrane computing. As a part of this study, we have explained some concepts of the p-systems. Concepts such as:

- Components
- Interactions between the components.
- The evolution of a p-system.

Moreover, we have focused our work on some concepts of the multiagent systems such as objects, agents, relations and resources. We have then established a link that connects both paradigms. Because of this we have proved that p-systems are a particular case of a multiagent system. With links such as the one we have built, we are able to describe any p-system in Multiagent system terms. Thus, the study of the p-systems can also be included in the study of the multiagent systems.

It is not intended to explain deeply neither the P-system technology nor the Multiagent system technology. There are many papers that have been published about that [Păun 1998], [Arroyo,2001] [Arroyo, 2003], The P-system described here has been generic. We have not explained into detail aspects as rules election, priorities between rules, etc. The general and main concepts have been useful to show our work. Again, MAS has been generally

described without focusing in every detail. In order to understand better the concepts of Multiagent system technology, refer to [Wooldridge,2002] "An Introduction to Multiagent Systems", Wiley, 2002. or [J.Ferber, 1999] "Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence". The Multiagent system described here has been generic.

We have built a "bridge" between these two universes and, because of that we have added some extra properties to the p-systems. The code provided here is a proposal but it is not the main goal of this work. By formally establishing an association between the P-systems and MAS we take a step on opening the p-systems to other technologies.

Proposals to create a link between P-systems and MAS exist already, Our proposal has been inspired in [Acampora, 2007]. However, our model has not been focused on the implementation but on providing new aspects to the current P-systems. In order to do this we have provided more details in the design of our mapping.

According to our association between p-systems and multiagent systems, a cell membrane can be treated as an independent and autonomous agent. By providing autonomy and independence to the cell membranes, we also change the perspective for the rest of the components within the p-systems. For example, when a membrane 'A' contains another membrane 'B', a dependency from 'B' to 'A' exists. This occurs because, whenever 'B' is dissolved, its objects become objects of the membrane 'A'. By looking at this scenario from a multi agent perspective, these two membranes are two agents that establish a relation. Because of this relation, the agent 'B' provides objects to the agent 'A'. This happens whenever the agent 'B' executes an action called 'dissolve.'

This perspective allows individual treatment of the p-systems components.

Nowadays we work with the p-system as an entire compacted block of components that are going through an evolutionary process. The p-system functioning is treated under a global perspective.

Under the MAS perspective, the analysis changes. The p-system works properly if and only if every agent  $a_1, a_2, \dots, a_n, \dots, a_{sync}$  works properly.

Thus, the analysis of the correct functioning of a given p-system is not based on the global processes but in the correction of all the agents which are part of the MAS. When every agent is assured of functioning correctly, we are assured that the entire multiagent system is working correctly.

By studying the characteristics of the p-system under the multi agent system perspective we can add the p-systems as a particular case of the current MAS.

This new perspective can also change the implementation of the p-systems. The use of the agent  $a_{sync}$  assures the synchronization of the entire multi agent system. This characteristic does not exist in the current implementation of p-systems.

Because the Multi agent systems are used broadly, they can be applied to many current systems. It is possible to state that, by using mappings such as this, we can reduce more systems into MAS. The key for this will be updating the rules in the MAS. When this occurs we would be able to reduce more systems into MAS and to assure correct functioning.

## Bibliography

---

- [Păun 1998] "Computing with Membranes", Journal of Computer and System Sciences, 61(2000), and Turku Center of Computer Science-TUCS Report n° 208, 1998.
- [Acampora, 2007] A Proposal of Multi-Agent Simulation System for Membrane Computing Devices Giovanni Acampora, Member, IEEE and Vincenzo Loia Member, IEEE 2007 IEEE Congress on Evolutionary Computation (CEC 2007).
- [Jimenez, 2006] Membranes as Multi-agent Systems: an Application to Dialogue Modelling Gemma Bel-Enguix and Dolores Jimenez Lopez Bel-Enguix. G., Lopez, D., T., 2006, in IFIP International Federation for Information Processing, Volume 218 Professional.
- [A. Arteta, 2008] "Algorithm for Application of Evolution Rules based on linear diophantine equations" Synasc 2008, Timisoara Romania September 2008 [1] A. Syropoulos, E.G. Mamas, P.C. Allilones, K.T. Sotiriades "A
- [Arroyo, 2001] "Structures and Bio-language to Simulate Transition P Systems on Digital Computers," Multiset Processing (.
- [Arroyo, 2003] "A Software Simulation of Transition P Systems in Haskell, Membrane Computing,"
- [Wooldridge, 2002] "An Introduction to MultiAgent Systems", Wiley, 2002.
- [Acampora, 2005] "Fuzzy control interoperability and scalability for adaptive domotic framework," in IEEE Transactions on Industrial Informatics, vol.1, no. 2, May 2005, pp. 97–111.
- [J.Ferber, 1999] "Multi-Agent Systems. An Introduction to Distributed Artificial Intelligence," Addison Wesley, 1999.

---

## Authors' Information

---

**Alberto Arteta Albert** – Associate professor U.P.M Crtra Valencia km 7, Madrid-28031, Spain;

e-mail: [arteta@eui.upm.es](mailto:arteta@eui.upm.es)

Research: Membrane computing, Education on Applied Mathematics and Informatics

**Angel Goñi** – Researcher natural computing group U.P.M, Madrid, Spain; e-mail: [anandgel@gmail.com](mailto:anandgel@gmail.com)

**Juan Castellanos** – Professor (UPM) Department Artificial Intelligence. Faculty of Informatics, Madrid 28660, Spain, e-mail: [jcastellanos@fi.upm.es](mailto:jcastellanos@fi.upm.es)

## IMPLEMENTATION OF GENETIC ALGORITHMS FOR TRANSIT POINTS ARRANGEMENT

Dmitry Panchenko, Maxim Shcherbakov

**Abstract:** The problem of transit points arrangement is presented in the paper. This issue is connected with accuracy of tariff distance calculation and it is the urgent problem at present. Was showed that standard method of tariff distance discovering is not optimal. The Genetic Algorithms are used in optimization problem resolution. The UML application class diagram and class content are showed. In the end the example of transit points arrangement is represented.

**Keywords:** transit points arrangement, genetic algorithms, optimization, software application.

**ACM Classification Keywords:** G.1.6 Optimization - Global optimization, D.1.1 Applicative (Functional) Programming

**Conference:** The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009

### Introduction

At present days the freight service billing and tariff distance defining are actual issues for railroad monopoly. There about 13000 railroad stations, wayside stops, passing-tracks and another points are located in Russia. We can compose the formal definition the set of these points connected with railway lines as a triply connected flow diagram. Problem for shortest distance discovering is very actual problem. So for the barest necessity there is a special reference book for freight service billing calculation between two various points of Russian Railroads. But there some problems are connected with search: it takes a lot of time and persons can make mistakes. Thus development fast search method and this implementation are very urgent problems.

### Standard method for tariff distance calculation

For contraction of adjacency matrix volume (columns count more than 169 000 000) using model of transit points (TP). In the reference book for freight service billing for each transit point  $P_i$ , is defined  $S_{ij}$  - shortest distance to nearest transit point  $TP_j$ , and distance between arbitrary points  $S_{jk}$ . This model allows to reduce information volume and complexity.

We should define distance between two arbitrary railroads points  $P_i$  and  $P_j$ . Let there are some points  $TP_k$  and  $TP_m$ . are located in the neighbourhood of  $P_i$ . Let there are some points  $TP_n$  and  $TP_q$ . are located in the neighbourhood of  $P_j$ . Using tables of departure and arrival we can define distances  $S_{ik}$ ,  $S_{im}$ ,  $S_{jn}$ ,  $S_{jq}$ . Using reference book for freight service billing we can define distances between TP:  $S_{kn}$ ,  $S_{kq}$ ,  $S_{mn}$  и  $S_{mq}$ . Thus we get the set of tracks  $L_i$ , where  $L_1 = S_{ik} + S_{kn} + S_{jn}$ ,  $L_2 = S_{ik} + S_{kq} + S_{jq}$ ,  $L_3 = S_{im} + S_{mn} + S_{jn}$ ,  $L_4 = S_{im} + S_{mq} + S_{jq}$ . Moreover tariff distance is defined as [Tarrif guide 1, 2001, Tariff guide 3, 2001]

$$L = \min_i L_i$$

In this case tariff distance is not shortest distance.

### GA method for shortest distance discovering

Let look at singular case. Let railroad points are located as showed at the Figure 1. In this case tariff distance between  $P_i$  and  $P_j$  will define as  $L_1 = S_{ik} + S_{kn} + S_{jn}$ . Obviously that tariff distance include  $S_{ik}$  two times. So optimal arrangement TP for new sectors of railroads is urgent problem, because it has an influence on inaccuracy in tariff distance calculation on the one hand and on the volume of reference book on the other hand.



Figure 1. Example of TP arrangement

For optimization problem solving we use the genetic algorithms (GA) [Batishev, 1995], because GA has a resistibility to hit into local minimum, and very effective on huge data set. Graph nodes which have two or more incoming connection are called main node. Generally transit points are located in the main nodes for optimal arrangement.

We define the chromosome as  $H=\{h_1, h_2, \dots, h_t\}$ , where  $t$  – main nodes count. At that

$$h_i = \begin{cases} 0, & \text{if } P_i \text{ is not } TP \\ 1, & P_i \text{ is } TP \end{cases}$$

Term “pass” is set of nodes and connections in the graph between two main nodes. We define number of passes with out TP as  $s_i$ ,  $r_{max}$  – highest possible number of TP in graph’s fragment is considered in the arrangement problem,  $r_i$  – number of TP in the  $i$  chromosome. In this case fitness function is calculates

$$f_i = \begin{cases} 0, & \text{if } r_i > r_{max} \\ \frac{1}{s_i}, & \text{if } r_i \leq r_{max} \end{cases}$$

After the individual fitnesses of the offspring has been evaluating we can define offspring selection probability for crossing. Offspring selection probability is directly proportional the individual fitnesses of the offspring. After that the crossover and mutation are executed.

Genetic algorithms using for arrangement.

Choose 2 individuals A and B with highest fitness

After crossover obtain individuals C.

Make the mutation of offspring C with probability **mutationPos**.

Make the inversion of offspring C with probability **inversionPos**.

Sort the set of chromosomes in the order of fitness decrease.

Change the the worst chromosome on C.

Input parameters are graph of rairod stations, maximum allowable number of TP, mutation and inversion probabilities and point of chromosomes crosses.

---

## Implementation

---

For implementation of Genetic Algorithms was realized software application. We use Microsoft Visual Studio Express Edition as a IDE and C# as a programming language and Test-Driven Development as a software development technology. UML application class diagram (without links) of the main classes is represented on the Figure 2. Class «Khromosome» includes necessary fields and methods (as the Inversion, Mutation, Crossover methods) for implement of individuals and offspring. Class «Generation» uses list of «Khromosome» class’ objects and contains methods for create and change of chromosomes list. The class «GeneticAlgorithm» uses object of «Generation» class and has initializing methods.

Bellow example of TP arrangement using software application is represented. The initial states map is showed at the Figure 3.

In this case after GA implementation for arrangement 2 transit points were added into the graph. They are showed at the Figure 4 and marked yellow points.

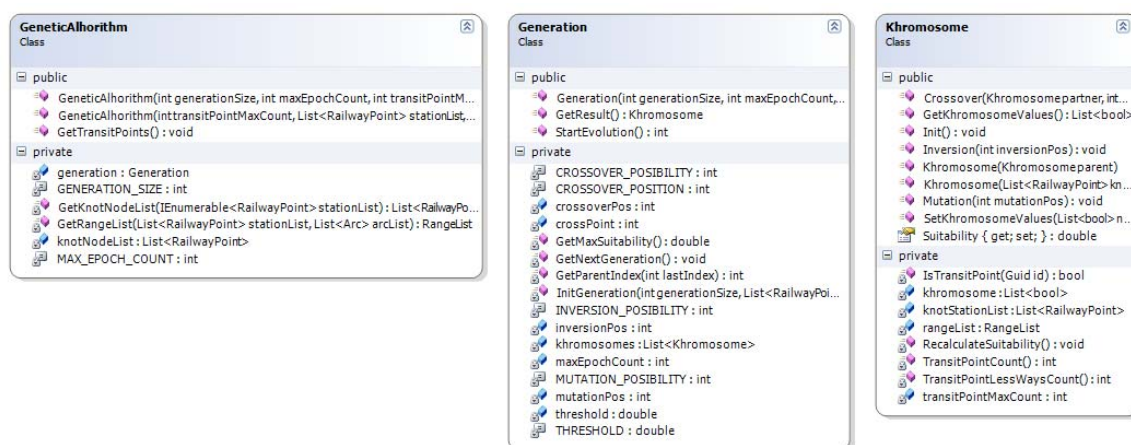


Figure 2. Class diagram for GA implementation.

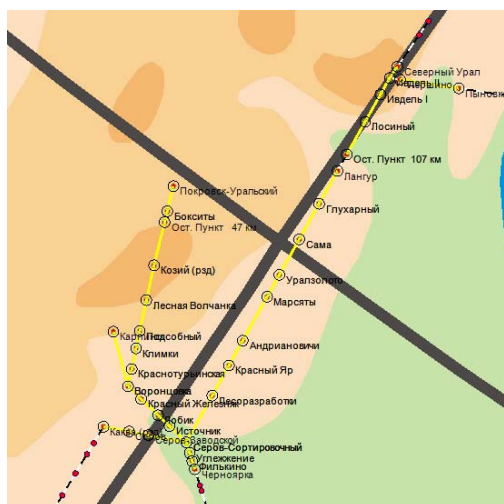


Figure 3. Initial railroad points map



Figure 4. Added transit points

## Conclusion

Offered method allows find optimal arrangement of transit points and as a result decrease inaccuracy in tariff distance calculation. Software application was tested on 15 various graphs with definite results for effectiveness and adequacy defining. We obtained the same results of methods implementation and the test samples with permissible error.

## Bibliography

- [Tariff guide 1, 2001] Tariff guide #4, Vol 1. Tariff distances between railroad station / Moscow, Transport, 2001  
 [Tariff guide 3, 2001] Tariff guide #4, Vol 1. Tariff distances between transit points / Moscow, Transport, 2001  
 [Batishev, 1995] Batishev D.I., Geneticheskie algoritmi reshenia ekstremal'nih zadach, Voronez, 1995

## Authors' Information

**Dmitry Panchenko** – Senior programmer Netvox Lab, Russia, e-mail: [pdmitry@netvoxlab.ru](mailto:pdmitry@netvoxlab.ru)

**Maxim Shcherbakov** – Assistant Professor; Volgograd State Technical University, Lenina Av., bl.28, Volgograd-400131, Russia; e-mail: [vstu@gebeus.ru](mailto:vstu@gebeus.ru)

## SELF EVOLVING CHARACTER RECOGNITION USING GENETIC OPERATORS

**Shashank Mathur**

**Abstract:** *In this paper, a novel approach for character recognition has been presented with the help of genetic operators which have evolved from biological genetics and help us to achieve highly accurate results. A genetic algorithm approach has been described in which the biological haploid chromosomes have been implemented using a single row bit pattern of 315 values which have been operated upon by various genetic operators. A set of characters are taken as an initial population from which various new generations of characters are generated with the help of selection, crossover and mutation. Variations of population of characters are evolved from which the fittest solution is found by subjecting the various populations to a new fitness function developed. The methodology works and reduces the dissimilarity coefficient found by the fitness function between the character to be recognized and members of the populations and on reaching threshold limit of the error found from dissimilarity, it recognizes the character. As the new population is being generated from the older population, traits are passed on from one generation to another. We present a methodology with the help of which we are able to achieve highly efficient character recognition.*

**Keywords:** *Genetic operators, character recognition, genetics, genetic algorithm.*

**ACM Classification Keywords:** *1.2 Artificial Intelligence, 1.4 Image processing and computer vision, 1.5 Pattern Recognition.*

**Conference:** *The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009*

---

### Introduction

Over the years, several methods have been proposed for offline character recognition which is the translation of characters provided as images into editable textual characters. Character recognition has been previously implemented for printed Tamil text [I], a traditional south Indian language. Some previously published works in character recognition such as mathematical expression recognition [II] and intelligent form based character recognition system [III] have been successful and have laid the stepping stone for further research in the area such as application of genetic operators to provide better results. Genetic operators derive their existence from Genetics, a branch of biology dealing with the heredity and variations in living organisms. The methodology presented here projects that new generations of solutions possess traits of their predecessors, thus implementing inheritance, a significant concept of genetics proposed by Johann Gregor Mendel [IV]. Genetic algorithms have been developed on these evolutionary concepts and applied in various fields such as knapsack problem [V], structural optimizations [VI], Assembly planning [VII] and Design of Large-Scale Reverse Logistic Networks in Europe's Automotive Industry [VIII].

With the establishment of evolutionary concepts and development of genetic programming, and works of various researchers such as Price who noticed that a covariance relationship exists between the number of successful offspring that an individual produces and the frequency of any given gene in that individual [IX], we are able to achieve a highly efficient evolutionary character recognition methodology with the help of genetic operators. In this paper, haploid chromosomes which have been implemented in the form of a bit pattern of dimensions 1x315 values, which can be operated upon by genetic operators like selection, crossover and mutation. An initial population is taken from which new generations of offspring's are obtained till the error is reduced till a threshold



limit and this method has been instrumental in the process of character recognition. A unique fitness function has been designed with the help of which the fittest set of offspring's are selected and ultimately lead us to the desired solution thus resulting in selection as a consequence of competition [X]. It has been previously seen that two parents result in two more offspring's, as done in evolutionary timetabling using biased genetic operators [XI].

In some previous works, Artificial Neural Networks have been used for optimizing both the architecture and the connection weights of multilayer feed forward neural networks [XII] and Arabic character recognition [XIII], but here the recognition process is carried out solely with the help of genetic operators making it a self evolving character recognition procedure. The methodology presented is significant in abridging the concepts of genetics and artificial intelligence, thereby introducing the various biological processes in programming and yielding efficient results. The various published works based on genetic algorithms like cryptanalysis method based on Genetic Algorithm and Tabu Search to break a Mono-Alphabetic Substitution Cipher in Adhoc networks[XIV] and the weight constrained shortest path problem[XV] have aided the development of the methodology presented his to ensure that results obtained from the evolutionary based algorithm provide high degree of efficiency.

The paper also presents the process of adaptation, an emergent property achieved by the Darwinian process of selection on heritable variation [XVI]. The new generations that are produced are adapted to the best solutions of the previous generations thereby providing us with a solution that is closer to the expected output as the newer generations are produced. The more the number of generations, more accurate the result is obtained as the error keeps reducing from one generation to another. The fittest solution survives at last which is provided as output. The methodology presented keeps knowledge about the fitness about all the individuals of the population of a particular generation which is lost on the development of a new generation. It is made sure that only a set of fit solutions are selected to generate a new generation so that an least error output is obtained in a optimum number of generations. Since only a set of fit values are considered to develop a new generation we get offspring's that have better fitness values as compared to the parents.

The purpose of this paper is to present new approach for character recognition based on the evolutionary theory with the application of genetic operators like selection, crossover, and mutation on sets of characters taken as an initial population. Bit patterns are used to represent haploid chromosomes which are operated upon by genetic operators and then used to generate new populations. With each new generation that is obtained, there is a decrease in the error between the pattern to be matched and the fittest offspring that was obtained in the last produced generation of solutions. The process of self evolution of recognized result from a population of patterns provides us with highly efficient results of character recognition.

---

## Methodology

---

The self evolving character recognition process is carried out with the help of bit patterns of size 1x315 values which have been obtained from monochromatic images of characters of size 21x15. A set of characters for each alphabet are converted into the form of single row bit patterns to generate an initial population from which new generations of offspring's are generated till a threshold limit of error between the character to be recognized and the fittest solution of a generation is obtained. The process of creation of new generations is suspended once this threshold limit is obtained. The genetic operators like selection, crossover and mutation have been described along with a unique fitness function with the help of which only a set of fit offspring's are selected and the others are discarded, thus implementing survival of the fittest.

### *A. Procedure*

The process of character recognition starts by reading an input monochromatic image, of any size which is resized to an image of 21x15 pixels and then converted into a single row bit pattern of dimension 1x315 which is then matched with the generated populations at various stages. A set of 130 characters, 5 characters each for all the 26 lower case alphabetical characters are considered as an initial population. The fitness function selected the set of fittest values to generate a new population which is then operated upon by the genetic operators to

yield new generation on which again the fitness function is applied to generate fit solutions, this process is carried out till a threshold limit for error between the input and fittest solution is obtained. In the process of character recognition it is essential that if the error is less than the threshold limit, it can be recognized with the help of lesser number of computations, thus making the algorithm provide us a result in an optimum number of computations, thereby making it cost efficient.

### *B. Fitness function*

This is a function with the help of which we are able to learn about the fitness of a solution, and whether that solution should be kept to generate a new population or not. With the help of this fitness function a set of offspring's are selected to generate a new generation of offspring's while the other values are discarded. We have haploid chromosomes in the form of single row bit pattern, thus in order to check whether a solution that has been generated should be kept for generating a new generation or should be discarded is based on the dissimilarity obtained between the offspring and the input. We now see the working of the fitness function, suppose we have two bit patterns, A, a single row bit pattern from the current population and B, single row bit pattern of input that is supposed to be recognized, then C is the obtained as the result of bitwise XOR of A and B, from which we obtain the dissimilarity coefficient, which is the number of '1' bits in the XOR result, which show us the number of bits for which the input and offspring differ.

A= 0 0 0 0 1 1 1 1 0 0 1 1 0 0

B= 0 1 0 0 1 0 1 1 0 0 0 1 0 0

C=A XOR B

C=0 1 0 0 0 1 0 0 0 0 1 0 0 0

Dissimilarity Coefficient = 3

As we see the result of the XOR operation in the above example, this is carried out for the entire population taken one at a time. The dissimilarity coefficients are saved and their mean is calculated. The members of the population with a dissimilarity coefficient of less than the mean value are considered fit to generate the next generation while the others are discarded.

### *C. Selection*

This is one of the genetic operators that are useful in selecting a set of population with the help of which a new population is generated. As explained above, dissimilarity coefficients are calculated for the entire population out of which only the members having a coefficient of more than mean value are selected and the others are not a part of the new population that is generated.

### *D. Crossover*

A genetic operator, which introduces inheritance of traits and adaptation in the new generations of offspring's. In crossover, two chromosomes overlap to provide two new offspring's which inherit the trait from their parents and are adapted towards a better fit solution. This biological crossover is implemented here with the help of two single row bit patterns which overlap at randomly varying lengths of 1 to 310 to generate new offspring's that have lower results of dissimilarity coefficients. Suppose we have two single row bit patterns, A and B of dimensions 1x315 each. Then C and D are two offspring's that are developed when the crossover takes place at a length of 5. The greater the extent of crossover, greater are the chances of getting a offspring which has a lower dissimilarity coefficient. The process of the crossover is shown in figure 1-3.

A= 0 0 0 0 1 1 ... 1 1 0 1 1 1 1 0

B= **0 1 1 0 1 0 ... 1 1 0 0 0 1 0 0**

C= 0 0 0 0 1 0 ... **1 1 0 0 0 1 0 0**

D=**0 1 1 0 1 1 ... 1 1 0 1 1 1 1 0**

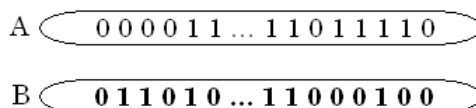


Fig. 1 Initial single row bit patterns.

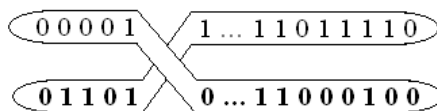


Fig. 2 Crossover at length of 5 bits.

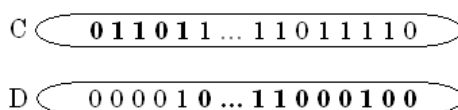


Fig. 3 Offspring's generated after crossover

### E. Mutation

This is the last genetic operator used. In case we have a very low dissimilarity coefficient of the order of less than 30 bits, which would be equivalent to an error of less than 10%, we do not populate new generations, but induce mutate the selected single row bit pattern at various lengths to generate a solution such that it is able to match our threshold limit for recognition of the character. Suppose we two single row bit patterns, A and B of dimensions  $1 \times 315$  each, where A is the input to be recognized and B is the pattern with less than 10% error then mutation is carried out. The process of mutation is shown below.

$$A = 10100000111$$

$$B = 10101010111$$

Dissimilarity Coefficient = 2

Few mutated results:

$$M1 = 10001010111$$

$$M2 = 10100010111$$

$$M3 = 10111010111$$

$$M4 = 10101010011$$

With the application of the induced mutations, we are able to generate a new population in which each offspring differs from the parent of the previous generation by only one bit. Thus we are able to achieve a solution in lesser number of computations thereby decreasing the overall complexity of the methodology presented.

### F. Algorithm

The methodology presented works according to the following algorithm shown in figure 4. We can observe that the working of the algorithm starts with the generation of an initial population. Each member of this initial population and the input image which is supposed to be recognized is applied to the fitness function from where we get the dissimilarity coefficients for each member of the population. If we get an error of less than 10%, we check if the error is less than 5%, if so then the member with the least dissimilarity coefficient is selected as the recognized. In case error is between 5% and 10%, such that character input which is quite similar to any of characters in the initial population we carry out mutation and generate a new population in which each member

differs from the input by only one bit. In case, we achieve an error of greater than 10%, we generate a new population by crossover of each member with the other members of the population at varying lengths. The new population thus generated is replaced with the initial population and we get iterations till the error reduces to 10%. The algorithm thus is capable of recognizing the characters with high efficiency.

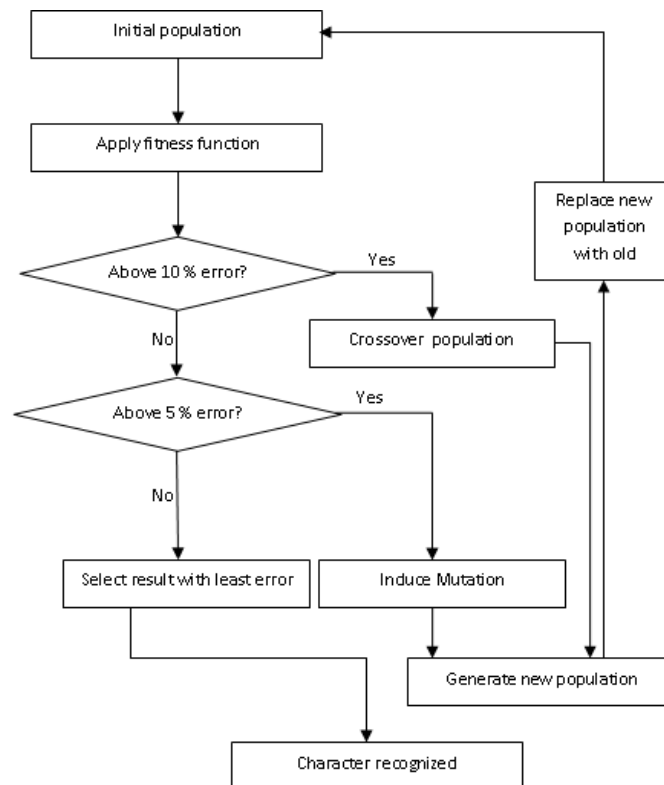


Fig. 4 Flowchart.

As the new generations are produced, the number of members of the population starts increasing. The number of individuals of a new population that is generated is found to be proportional to the number of individuals selected to generate it. The character recognition process carried out here has been implemented using a small initial population of 130 members which grows to form populations of hundreds and thousands in successive generations and leads us to find an optimum solution.

## Results

The self evolving character recognition using genetic operators was tested for 10 character inputs for each alphabet, thus a total of 260 samples and observed that a high efficiency of recognition is obtained as shown in the graph in figure 5. An overall efficiency of 79.23% was obtained for the character recognition process during the testing process. The methodology presented here was implemented on a Pentium 4 (3.4 GHz), 2GB RAM and MATLAB 7.0. We can note that the number of generations that are being created increase with the increase in dissimilarity coefficient. A character with an error of less than 5% is recognized with the help of the member of population having the least dissimilarity coefficient. Along with the genetic operators, the fitness function has been instrumental in providing a good estimate of the dissimilarity coefficient with the help of which unwanted members of the population can be discarded. Thus successfully we are able to implement concepts of genetics, a discipline of biology in character recognition process yielding high rate of efficiency.

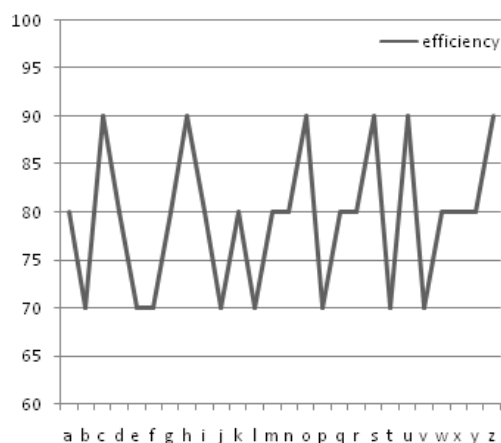


Fig. 5 Graph obtained for testing of character recognition method

## Conclusion

The paper has presented a new method of character recognition using a unique and robust methodology by the application of genetic operators. The paper is significant in abridging the concepts of genetics and artificial intelligence and thus providing a stepping stone for further research in this area. The highly efficient character recognition results reflect the accuracy of the implementation of the genetic operators. The innovative fitness function used here provides us with a dissimilarity coefficient which restrains the entry of unwanted members of the population which would only increase the computational cost of the methodology presented. Character recognition task is indeed a tough task which can be carried out by the methodology presented in a biologically inspired self evolving manner implementing all the laws of genetics. The paper has provided a link of evolutionary sciences in the field of computer science.

## Acknowledgement

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

## Bibliography

- [I.] Abnikant Singh, Markandey Singh, Ritwaj Ratan, Sarvesh Kumar, "Optical Character Recognition for printed Tamil text using Unicode", Journal of Zhejiang University, pp. 1297-1305, 2005.
- [II.] Erik G. Miller, Paul A. Viola, "Ambiguity and Constraint in Mathematical Expression Recognition", American Association for Artificial Intelligence, 1998.
- [III.] Dipti Deodhare, NNR Ranga Suri R. Amit, "Preprocessing and Image Enhancement Algorithms for a Form-based Intelligent Character Recognition System", International Journal of Computer Science and Applications Vol. II, No. II, pp.131-144, 2005.
- [IV.] Weiling F, "Historical study: Johann Gregor Mendel 1822-1884", American Journal of Medical Genetics, Vol. 40, No.1, pp.1-25, 1991.
- [V.] P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, "A monte carlo study of genetic algorithm initial population generation methods", Proceedings of the Winter Simulation Conference, pp. 543-547, 1999.
- [VI.] Alonso J. Juvinao Carbone, Ivan F. M. Menezes, Luiz Fernando Martha, "Mooring Pattern Optimization using Genetic Algorithms", 6th World Congresses of Structural and Multidisciplinary Optimization, Brazil, 2005.
- [VII.] Shana Shiang-Fong Smith (Shiang-Fong Chen), Yong-Jin Liu, "The Application of Multi-Level Genetic Algorithms in Assembly Planning", Journal of Industrial Technology, Volume 17, Number 4, pp.1-9. 2001.

- [VIII.] Ralf Schleiffer, Jens Wollenweber, Hans-Juergen Sebastian, Florian Golm, Natasha Kapoustina, "Application of Genetic Algorithms for the Design of Large-Scale Reverse Logistic Networks in Europe's Automotive Industry" Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.
- [IX.] Jeffrey K. Bassett, Mitchell A. Potter, Kenneth A. De Jong, "Applying Price's Equation to Survival Selection" Genetic and Evolutionary Computation Conference, USA, pp. 1371-1378, 2005.
- [X.] D. B. Fogel and J. W. Atmarz "Comparing Genetic Operators with Gaussian Mutations in Simulated Evolutionary Processes Using Linear Systems", Biological Cybernetics., pp. 111-114, 1990.
- [XI.] Daniel Danciu, "Evolutionary Timetabling Using Biased Genetic Operators" Journal of Computing and Information Technology, pp. 193-199, 2003.
- [XII.] Morgan Kaufmann, "Genetic Programming of Minimal Neural Nets Using Occam's Razor", Proceedings of 5<sup>th</sup> international conference on Genetic Algorithms, pp. 342-349, 1993.
- [XIII.] Ahmad M. Sarhan, Omar I. Al Helalat, "Arabic character recognition using artificial neural networks and statistical analysis", Proceedings of world academy of science, engineering and technology, Vol. 21, pp. 32-36, 2007
- [XIV.] K. Verma, Mayank Dave, R. C. Joshi, "Genetic Algorithm and Tabu Search Attack on the Mono-Alphabetic Substitution Cipher in Adhoc Networks", Journal of Computer Science, Vol. 3, No. 3, pp. 134-137, 2007.
- [XV.] Jeffrey K. Bassett, Mitchell A. Potter, Kenneth A. De Jong, "Applying Price's Equation to Survival Selection" Genetic and Evolutionary Computation Conference, USA, pp. 1371-1378, 2005.
- [XVI.] Anthony V. Sebald and Lawrence J. Fogel, "Emergent phenomena in genetic programming", in Proceedings of the Third Annual Conference on Evolutionary Programming, \_ pp. 233-241, 1994.
- 

### Authors' Information

---

***Shashank Narain Mathur*** - Bachelors of Technology (Computer Science Engineering), Amity School of Engineering and Technology affiliated to Guru Gobind Singh Indraprastha University, New Delhi, India.  
e-mail: [shashanknarainmathur@yahoo.com](mailto:shashanknarainmathur@yahoo.com)

## GENERATING MORE BOUNDARY ELEMENTS OF SUBSET PROJECTIONS

Hasmik Sahakyan, Levon Aslanyan

**Abstract:** Composition problem is considered for partition constrained vertex subsets of  $n$  dimensional unit cube  $E^n$ . Generating numerical characteristics of  $E^n$  subsets partitions is considered by means of the same characteristics in  $n-1$  dimensional unit cube, and construction of corresponding subsets is given for a special particular case. Using pairs of lower layer characteristic vectors for  $E^{n-1}$  more characteristic vectors for  $E^n$  are composed which are boundary from one side, and which take part in practical recognition of validness of a given candidate vector of partitions.

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

**ACM Classification Keywords:** G.2.1 Discrete mathematics: Combinatorics

**Conference:** The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009

---

### Introduction

---

Set systems are considered represented as subsets of vertices of  $n$  dimensional unit cube  $E^n$ . Given a subset of  $E^n$ , its numerical characterization is composed by partitions and their sizes, which are the coordinates of the corresponding associated vector of partitions. General numerical characterization of vertex subsets of  $E^n$  through their partitions is considered in [SA, 2006], [S, 2009] where a complete and simple structural description of the set of all integer-valued vectors, which serve as associated vectors of partitions is given, The description is through the set of its boundary elements. Elements of boundary set with minimum and maximum weight are known by [SA 2001], [SA, 2006]. Current research focuses on generating more boundary elements via the known ones for smaller dimensions, and constructs a set of corresponding subsets of vertices.

---

### Basic Structural Description

---

Let  $E^n$  be the set of vertices of  $n$ -dimensional unit cube,  $E^n = \{(x_1, \dots, x_n) / x_i \in \{0,1\}, i = 1, \dots, n\}$ . For an arbitrary variable  $x_i$ , consider partition of the cube into two subcubes according to the value of  $x_i, 1 \leq i \leq n$ . Denote these subcubes by  $E_{x_i=1}^{n-1}$  and  $E_{x_i=0}^{n-1}$  correspondingly. Similar to this, each subset of vertices  $M \subseteq E^n$  can be partitioned into  $M_{x_i=1}$  and  $M_{x_i=0}$ . For a given  $m, 0 \leq m \leq 2^n$  consider an  $m$ -element subset  $M$  of  $E^n$ . Vector  $S = (s_1, \dots, s_n)$  is called associated vector of partitions of  $M$  if  $s_i = |M_{x_i=1}|$  for all  $1 \leq i \leq n$ . Let  $\psi_m(n)$  denote the set of all associated vectors of partitions of  $m$ -subsets of  $E^n$ . Let  $\Xi_{m+1}^n$  denote the  $n$  dimensional  $(m+1)$ -valued grid, i.e., the set of all integer-valued vectors  $S = (s_1, s_2, \dots, s_n)$  with  $0 \leq s_i \leq m, i = 1, \dots, n$ . Obviously  $\psi_m(n) \subseteq \Xi_{m+1}^n$ .

Vector  $S \in \psi_m(n)$  is called an upper (lower) boundary vector for  $\psi_m(n)$  if no vector  $R \in \Xi_{m+1}^n$  with  $R > S$  ( $S > R$ ) belongs to  $\psi_m(n)$ . By  $\widehat{\psi}_m(n)$  and  $\widetilde{\psi}_m(n)$ , respectively, denote the sets of all upper and lower boundary vectors of  $\psi_m(n)$ . The sets  $\widehat{\psi}_m(n)$  and  $\widetilde{\psi}_m(n)$  contain equal numbers of elements [S, 2009]. Let  $r$  denote this number of elements and let  $\widehat{\psi}_m(n) = \{\widehat{S}^1, \dots, \widehat{S}^r\}$  and  $\widetilde{\psi}_m(n) = \{\widetilde{S}^1, \dots, \widetilde{S}^r\}$  so that  $(\widehat{S}^j, \widetilde{S}^j)$  is a pair of complementary vertices:  $\widehat{S}_i^j = m - \widetilde{S}_i^j$ , for  $1 \leq i \leq n$ . For each vector  $\widehat{S}^j$  all coordinates are greater than or equal to  $m/2$  and for each vector  $\widetilde{S}^j$  all coordinates are less than or equal to  $m/2$ . For a pair  $\widehat{S}_j, \widetilde{S}_j$ , from  $\widehat{\psi}_m(n)$  and  $\widetilde{\psi}_m(n)$  respectively, let  $I(\widehat{S}_j)$  (the notion  $I(\widetilde{S}_j)$  also may be used) denotes the vector set of the sub-cube of  $\Xi_{m+1}^n$  spanned by these vectors:  $I(\widehat{S}_j) = \{Q \in \Xi_{m+1}^n / \widehat{S}_j \leq Q \leq \widetilde{S}_j\}$ . proves that the collection of all sub-cubes  $\{I(\widehat{S}_j) / \widehat{S}_j \in \widehat{\psi}_m(n)\}$  composes the basic set  $\psi_m(n)$ :

$$\text{Theorem 1. } \psi_m(n) = \bigcup_{j=1}^r I(\widehat{S}_j)$$

## Boundary Cases

Basic object of analysis in this section is the set of all monotone Boolean functions in  $E^n$ , with exactly  $m$  ones. Let  $M_m^1(n)$  be the set of all associated vectors of partitions of  $m$ -subsets corresponding to the ones of these monotone Boolean functions. Similarly  $M_m^0(n)$  is the set of all associated vectors of partitions of  $m$ -subsets corresponding to the zeros of such monotone Boolean functions. It is easy to check that  $\widehat{\psi}_m(n) \subseteq M_m^1(n)$  and that  $\widetilde{\psi}_m(n) \subseteq M_m^0(n)$ . Thus for description of  $\psi_m(n)$  suffice to find those monotone Boolean functions which correspond to  $\widehat{\psi}_m(n)$  (and/or  $\widetilde{\psi}_m(n)$ ). Suppose the weights (sum of all coordinates) of vectors of  $\widehat{\psi}_m(n)$  belong to some interval  $[L_{\min}, L_{\max}]$ . A specific set of monotone Boolean functions is constructed in [AS, 2001], for which the corresponding associated vectors of partitions belong to  $\widehat{\psi}_m(n)$  and have the weight  $L_{\min}$ . The value of  $L_{\min}$  and the coordinates of associated vectors are also analyzed in detail.

Let  $D^{i_1, \dots, i_n}$  be the set of vertices of  $E^n$  arranged in decreasing order of numeric value of the binary vectors  $\langle x_{i_1}, \dots, x_{i_n} \rangle$ . It is easy to check that the first  $2^{n-1}$  elements of  $D^{i_1, \dots, i_n}$  form the set of vertices of  $E_{x_{i_1}=1}^{n-1}$  and the remainder  $2^{n-1}$  elements form  $E_{x_{i_1}=0}^{n-1}$  being arranged in the same decreasing order of numeric value of  $\langle x_{i_2}, \dots, x_{i_n} \rangle$ . Denote these sets by  $D_{x_{i_1}=1}^{i_2, \dots, i_n}$  and  $D_{x_{i_1}=0}^{i_2, \dots, i_n}$  respectively. Similarly the first  $2^{n-2}$  elements of  $D_{x_{i_1}=1}^{i_2, \dots, i_n}$  form  $E_{x_{i_1}=1, x_{i_2}=1}^{n-2}$  and the rest  $2^{n-2}$  elements form  $E_{x_{i_1}=1, x_{i_2}=0}^{n-2}$  - arranged in decreasing order of numeric value of the vectors  $\langle x_{i_3}, \dots, x_{i_n} \rangle$ , etc. It follows that each initial part (subset) of  $D^{i_1, \dots, i_n}$  (respectively  $D_{x_{i_1}=1}^{i_2, \dots, i_n}$ ,  $D_{x_{i_1}=0}^{i_2, \dots, i_n}$ , etc.) of arbitrary size serves as the set of ones of some monotone Boolean function. Let  $D^{i_1, \dots, i_n}(m)$  denote the  $m$ -length initial segment of  $D^{i_1, \dots, i_n}$ . Denote by  $D^n$  the overall set of all enumerations  $D^{i_1, \dots, i_n}$ , and let  $D^n(m)$  be the  $m$ -length initial segments of enumerations of  $D^n(m)$ .



The class of monotone Boolean functions, which have sets of ones belonging to  $D^n(m)$  we denote by  $f^{D^n(m)}$  and let  $S^{D^n(m)}$  be the corresponding set of associated vectors of partitions.

**Theorem 2.**  $S \in \widehat{\psi}_m(n)$  has weight  $L_{\min}$  if and only if  $S \in S^{D^n(m)}$ .

The proof of theorem which is not the target to bring here is by induction on  $n$ .

**Note.** Let us consider vector  $S = (s_1, \dots, s_n)$  of  $S^{D^n(m)}$ , which is the associated vector of partitions for some  $D^{i_1, \dots, i_n}(m)$ . It is easy to check that  $s_{i_1} \geq s_{i_2} \geq \dots \geq s_{i_n}$ . Further, vector  $S$  obeys the following very useful property: any vector  $(s_{i_1}, \dots, s_{i_j} + 1, *, \dots, *)$  for arbitrary  $j, 1 \leq j \leq n$ , does not belong to  $\psi_m(n)$ .

---

### Generating New Boundary Elements

---

Thus, we consider the problem of description of all monotone Boolean functions, which correspond to  $\widehat{\psi}_m(n)$ . Let we have the solution of this problem for  $n-1$ : given  $m_1, 0 \leq m_1 \leq 2^{n-1}$ , and we may describe the monotone Boolean functions corresponding to  $\widehat{\psi}_{m_1}(n-1)$ .

Consider the way of composing the target construction through composing the ones which we have in dimensions  $n-1$ : consider partition of  $E^n$  according to some variable  $x_i$ , and consider the pairs of monotone Boolean functions in  $E_{x_i=1}^{n-1}$  and  $E_{x_i=0}^{n-1}$  respectively, which correspond to the upper boundary vectors and find such pairs, that in  $E^n$  (their union in  $E^n$ ) correspond to  $\widehat{\psi}_m(n)$ .

Consider partition of  $E^n$  according to the value of  $x_1$  and consider monotone Boolean functions in  $E_{x_1=1}^{n-1}$  and  $E_{x_1=0}^{n-1}$  with  $m_1$  and  $m_2$  ones respectively, where  $m_1, m_2$  are arbitrary partition of  $m$  with the only requirement that  $m_1 \geq m_2$ . We intend to find feasible pairs of monotone functions from  $E_{x_1=1}^{n-1}$  and  $E_{x_1=0}^{n-1}$  to get at first monotone functions in  $E^n$ , and then - monotone functions which correspond to  $\widehat{\psi}_m(n)$ .

Consider the following particular case.

In  $E_{x_1=1}^{n-1}$ , as well as in  $E_{x_1=0}^{n-1}$  consider monotone functions for which corresponding associated vectors belong to sets  $\widehat{\psi}_{m_1}(n-1)$  and  $\widehat{\psi}_{m_2}(n-1)$  respectively, and have the minimum possible weights. By theorem 2 these are functions of the form  $f^{D^{n-1}(m_1)}$  and  $f^{D^{n-1}(m_2)}$  respectively. Consider the case of the same order of variables, let it be  $D^{2, \dots, n}(m_1)$  and  $D^{2, \dots, n}(m_2)$ . Denote by  $S' = (s'_2, \dots, s'_n)$  and  $S'' = (s''_2, \dots, s''_n)$  - the corresponding associated vectors. Obviously this pair is feasible for getting a monotone function in  $E^n$ . Now we prove that it is feasible for getting an upper boundary vector as well, that is the vector  $S = (m_1, s'_2 + s''_2, \dots, s'_n + s''_n) = (s_1, s_2, \dots, s_n)$  belongs to  $\widehat{\psi}_m(n)$ .

The choice of this particular case is caused by theorem 2; that is that we have the description of monotone Boolean functions which correspond to the upper boundary vectors of minimum weight.

**Theorem 3.**  $S = (m_1, s'_2 + s''_2, \dots, s'_n + s''_n) = (s_1, s_2, \dots, s_n) \in \widehat{\psi}_m(n)$

Proof.

Suppose this is not the case and the vector  $\tilde{S} = (\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_n)$ ,  $\tilde{S} > S$  serves as the associated vector of partitions for some monotone Boolean function (component wise comparison is used).

Consider cases:

a)  $\tilde{s}_1 = s_1$ . Consider partition of  $E^n$  according to the value of  $x_1$ . We get monotone functions in  $E_{x_1=1}^{n-1}$  and  $E_{x_1=0}^{n-1}$  with  $s_1$  and  $m - (s_1 + 1)$  ones respectively. Denote by  $(s_2^*, \dots, s_n^*)$  and  $(s_2^{**}, \dots, s_n^{**})$  the corresponding associated vectors. Let  $\tilde{s}_2 = s_2, \dots, \tilde{s}_{i-1} = s_{i-1}$  and  $\tilde{s}_i > s_i$ . We confirm that  $s_2^* = s_2'$  and  $s_2^{**} = s_2''$ ,  $\dots$ ,  $s_{i-1}^* = s_{i-1}'$  and  $s_{i-1}^{**} = s_{i-1}''$ , otherwise  $s_2^* > s_2'$  (similarly  $s_3^* > s_3'$ ,  $\dots$ ,  $s_{i-1}^* = s_{i-1}'$ ) implies  $S' = (s_2', \dots, s_n') \notin \widehat{\psi}_{m_1}(n-1)$  and  $s_2^{**} > s_2''$  (similarly  $s_3^{**} > s_3''$ ,  $\dots$ ,  $s_{i-1}^{**} = s_{i-1}''$ ) implies  $S'' = (s_2'', \dots, s_n'') \notin \widehat{\psi}_{m_2}(n-1)$  due to the note above.  $\tilde{s}_i > s_i$  implies either  $s_i^* > s_i'$  or  $s_i^{**} > s_i''$ . By the same reasoning each of them leads to contradiction.

b)  $\tilde{s}_1 > s_1$ . Consider partition of  $E^n$  according to the value of  $x_1$ . We get monotone functions in  $E_{x_1=1}^{n-1}$  and  $E_{x_1=0}^{n-1}$  with  $\tilde{s}_1$  and  $m - \tilde{s}_1$  ones respectively. There exist  $\tilde{s}_1 - (m - \tilde{s}_1)$  vertices in  $E_{x_1=1}^{n-1}$  belonging to the set of ones of the function, for which the corresponding vertices in  $E_{x_1=0}^{n-1}$  does not belong to the set of ones of the function. We move these vertices from  $E_{x_1=1}^{n-1}$  to  $E_{x_1=0}^{n-1}$ . It follows from the case a) that  $\tilde{s}_2 = s_2, \dots, \tilde{s}_n = s_n$ . To provide  $s_2$  in  $E^n$  we need to provide  $s_2'$  in  $E_{x_1=1}^{n-1}$  and  $s_2''$  in  $E_{x_1=0}^{n-1}$ , otherwise either  $S' = (s_2', \dots, s_n') \notin \widehat{\psi}_{m_1}(n-1)$  or  $S'' = (s_2'', \dots, s_n'') \notin \widehat{\psi}_{m_2}(n-1)$ , etc. By theorem 2 the obtained sets are  $D^{2, \dots, n}(m_1)$  and  $D^{2, \dots, n}(m_2)$  respectively, which is contradiction since  $m_1 \geq m_2$ .

As a corollary we get that for any  $\tilde{m}$ ,  $m/2 \leq \tilde{m} \leq m$ , a vector of  $\widehat{\psi}_m(n)$  exists with a coordinate which equal to  $\tilde{m}$ .

We may also mention the topic of construction of  $m_1$ -subsets of vertices with the given associated vector of partitions  $S = (m_1, s_2' + s_2'', \dots, s_n' + s_n'') = (s_1, s_2, \dots, s_n)$ .

Given that the sets  $D^{2, \dots, n}(m_1)$  and  $D^{2, \dots, n}(m_2)$  can be constructed by the “interval bisection” method [SA, 2006], then a construction of the required set can be provided in the following way: construct first column putting  $m_1$  consecutive ones and  $(m - m_1)$  consecutive zeros; continue the construction in this 2 sections by the same “interval bisection” method.

The weight of  $S$  and the values of coordinates can be calculated.

---

## Conclusion

Generating numerical characteristics of partitions by means of the same characteristics in  $n - 1$  dimensional unit cube for a particular case, leads to more boundary elements. The corresponding subsets can be easily constructed.

---

**Acknowledgement**

---

The paper is partially financed by the project **ITHEA XXI** of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria ([www.ithea.org](http://www.ithea.org), [www.foibg.com](http://www.foibg.com)).

---

**Bibliography**

---

- [R, 1963] H. J. Ryser. Combinatorial Mathematics, The Mathematical Association of America, 1963, p. 150.
- [A, 1979] L. Aslanyan. Isoperimetry problem and related extremal problems of discrete spaces, Problemy Kibernetiki, 36, pp. 85-126 (1976).
- [GGP 1997] R.J.Gardner, P. Gritzmann, D. Prangenberg. On the computational complexity of reconstructing lattice sets from the X-rays. Technical Report (970-05012), Techn.Univ. Munchen, 1997.
- [AS, 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.
- [SA, 2006] H. Sahakyan, L. Aslanyan. The boundary descriptors of the n-dimensional unit cube subset partitioning, Journal "Information Theories and Applications", ISSN 1310-0513, Volume 13 (2006), Number 3.
- [S, 2009] H. Sahakyan. Numeral characterization of n-cube subset partitioning, Discrete Applied Mathematics, vol.157 (2009), issue 9, 2191-2197.

---

**Authors' Information**

---

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

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



---

## **METASPEED: Metadata ExTraction for Automatic SPEcifications of E-Documents**

*Science-research Project: Automated Metadata Generating for e-Documents Specifications and Standards*

*Supported by: **Bulgarian National Science Fund** under contract **D002-308***

*Thematic priority: 4. Information and Communication Technologies*

*Duration: 2009-2011*

---

### **What is the METASPEED Project**

**Metadata ExTraction for Automatic SPEcifications of E-Documents – METASPEED** is a Bulgarian research project funded by the Bulgarian National Science Fund under the thematic priority: Information and Communication Technologies. It aims to facilitate the development of Bulgarian standards and even commonly accepted specifications for the description of metadata for e-documents in different subject areas. Project partners include Bulgarian researchers from state and private universities and Bulgarian Academy of Sciences. This project is carried out by a consortium composed of: University of Plovdiv, Technical University of Sofia, New Bulgarian University and Institute of Mathematics and Informatics, Bulgarian Academy of Science.

The goal of this project can be briefly summarized as follows: to investigate and create technologies, methods and tools for automatic generation of metadata thus facilitating the proper specification of documents with different e-format, content and location. The rationale behind this goal is that e-documents are to be described by common schemas and rules for the purpose of retrieval, sharing and using. Usually documents in digital repositories are determined according to a particular specification and/or standard together with data about the document itself, i.e. metadata. As a rule the application of any standard requires too much metadata that are produced by experts in the subject area. Thus building an object repository appears to be a very labor consuming activity carried out by highly qualified people. Consider the electronic resources e.g. tests, learning content, etc. in the National Educational Portal (<http://start.e-edu.bg/>). These resources obey no unique standard. That is why our research efforts towards a standardization and automatic generation of metadata for different format e-documents are an economically motivated activity. Project findings will facilitate the access to different digital collections in a straightforward manner. This is the first stage toward the development of a uniform information environment in Bulgaria. We expect that the main contributions include:





- development of proper tools for an automatic metadata generation for collections containing digital documents of different shapes and types;
- building a framework to share European and Bulgarian e-resources;
- development of national standards for document sharing.

---

### **Partners**

The METASPEED project is an interdisciplinary project. This justifies the participations of people interested in computer linguistics, e-learning, standards for e-documents, multimedia applications, archival sciences, database systems, etc. The partners and their competences are presented in Table 1:

Table 1 Partners of the METASPEED Project

	PARTNER	COMPETENCES
	University of Plovdiv Department of Computer Informatics	e-learning ,computer linguistic
	Institute of Mathematics and Informatics - BAS Department of Information Systems	analysis, synthesis and retrieval of structured data from texts, images and video
	New Bulgarian University Department of Informatics	cognitive sciences, database systems, e-learning
	Technical University of Sofia Research laboratory "Technologies and Standards for e-Learning"	standards and systems for e-learning

### Project Work Packages

The project is built up of four work packages. Certainly significant dissemination and valorization activities are foreseen.

#### ***WP1. Standards of e-documents and tools for their automatic generation***

The goal of this package is to finalize the research analysis in the area and to prepare state-of the-art reports concerning:

- a) standards for e-learning;
- b) standards for cultural heritage multimedia documents;
- c) prescriptions for Bulgarian standards in different subject areas;
- d) methods for automatic metadata generation.

It is expected that project technical prescriptions of national standards for e-learning and specifications for cultural heritage e-objects will be proposed. For example the adoption or development of a standard, in the field of e-learning, could provide sharing (including export/import), multiplication and adaptation of learning resources (courses, materials and tests) hosted in Internet. During the adaptation of well-known standards and specifications a reasonable question arises: to what extent the national specifics such as language, educational system and traditions are to be considered.

#### ***WP2. Automated metadata generation from text documents***

The goal of WP2 is to develop methods, algorithms and tools to retrieve structured data from electronic text documents, written in different languages taking into account existing standards and specifications. To realize this goal a review-analysis of the existing methods and algorithms in the world and in particular – for Bulgarian language will be carried out..

---

The following tasks will be performed:

- critical analyses of existing methods and tools for retrieval of metadata from electronic texts;
- investigation of specialized technologies and methods for metadata retrieval for documents in different areas;
- Design and implementation of proper software prototypes;
- Experiments with the realized methods and tools for metadata retrieval on the documents in examined areas.

---

### ***WP3. Automated metadata generation from multimedia documents***

---

In the next three years, the world will create more data than has been produced in all of human history. It is well known that the search power of current searching engines is typically limited to text and its similarity, since less than 1% of the Web data is in textual form, the rest being of multimedia/streaming nature, particularly since a large portion of pictures still remains as "unstructured data". The Enterprise Strategy Group (<http://www.enterprisestrategygroup.com/management>) estimates that more than 80 milliard photographs are taken each year. Using of digital images promises to emerge as a major issue in many areas, for instance Google answers daily more than 200 million queries against over 30 milliard items, because of this we need to extend our next-generation search to accommodate these heterogeneous media.

Some of the current engines search the data types according to textual information or other attributes associated with the files. An orthogonal approach is the Content-based Image Retrieval (CBIR). It is not a new area – in current surveys can be counted more than 300 systems, most of them exemplified by prototype implementations. The typical database size is in the order of thousands of images - very recent publicly-available systems, such as ImBrowse (<http://media-vibrance.itn.liu.se/>), Tiltomo (<http://www.tiltomo.com/>) and Alipr (<http://www.alipr.com/>), declare to index hundreds of thousands of images.

The user questions in image search are partitioned into three main levels:

- *Low level* – this level includes basic perceptual features of visual content (dominant colors, color distribution, texture pattern, etc.);
- *Intermediate level* – this level forms next step of extraction from visual content, connected with emotional perceiving of the images, which usually is difficult to express in rational and textual terms. The visual art is area, where these features play significant role. Typical features in this level are color contrasts, because one of the goals of the painting is to produce specific psychological effects in the observer, which are achieved with different arrangements of colors;
- *High level* – this level includes queries according to rational criterions. In many cases the image itself does not contain information which would be sufficient to extract some of the characteristics. For this reason current high-level semantic systems still use huge amount of manual annotation.

Usually the existing systems for image retrieval are limited by the fact that they can operate only at the primitive feature level, while users operate at a higher semantic level. This mismatch is often referred as *semantic gap*. The Project aims at finding and analyzing new content-based image retrieval methods to analyze, index, and retrieve the image and video. The goal is to increase the retrieval effectiveness by a proper choice of image features from the MPEG-7 standard and on that base finding the description of some concept, which humans use in their everyday life.

---

**WP4. Automated creating and testing of digital repositories in different areas: A) cultural heritage; B) e-learning; C) spatial information systems; D) automated referring of scientific publications.**

---

The main task is creating and studying of methods and tools for *automated metadata generating from Internet-pages (in corresponded thematic domains)*. There will be examined and adapted known technologies for search in Internet-pages, using tools for automated metadata generating from text and multimedia. As a result Internet-space will be used as a source for creating of digital repositories and for testing corresponded methods and tools. Other digitalized sources for carrying out the experiments are created during the years collections of electronic resources of the Project partners in the field of e-learning and scientific publications.

Basic interest during leading the *investigations of p.A* will be separate on studying and elaborating of search methods in document collections (contained text and graphic objects) and their automated classifying in appropriate ontological systems.

In *investigations of p.B* special attention will be focused on the question for *automated metadata retrieval* from learning resources (for instance: concepts, relations between them, degree of complexity, etc.) and generating of new learning objects (for instance test tasks and e-courses), rendering an account the results of leading e-learning (for different types of learners). For this purpose, except leading a linguistic analysis (processing of the content of used learning materials and tests), it is necessary (in order to data extraction for so cold *portfolio* of each learner) to trace and to analyze dialogue between learners and lecturers, learners and system for content management, etc.

The INSPIRE standards for spatial meta-data are obligatory for the European Union member states. This means that it is very important the cultural space information to be described following these standards. Two main tasks are connected to this problem:

- to create Bulgarian standards and thesauri for spatial meta-data of the cultural heritage objects, which are corresponding to the INSPIRE standards;
- following these standards, to develop methods and tools for meta-data extraction from the cultural objects' descriptions.

In case of *investigations in p.D* it is posed a question for creating repository of scientific publications review (using automated metadata retrieval from full-text articles). Referring of scientific articles compels referent-experts to review a lot of articles in journals and books, connected with given scientific area, to classify and to summarize on thematic directions. Main problem of the referring is a quantity of work, which referents have to do, to find important for the domain articles and results. Automating of the preliminary selection and classification of the incoming papers can be assist referring and decrease the probability for gaps. In the same time, automated solving of the opposite task – how is adequate one review to the content of the corresponding paper – is also important.

Automated metadata retrieval from the collection of scientific publications in different languages (in the case of specialized texts in concrete scientific domain) and theirs classifying in corresponded ontological system is important in the case of multilingual referring journals (for instance – containing reviews in English, Russian and Bulgarian languages) increase the quality of the issue.



