

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin  
(editors)

**Information Models  
of  
Knowledge**

**ITHEA<sup>®</sup>  
KIEV – SOFIA  
2010**

**Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)**

**Information Models of Knowledge**

ITHEA®

Kiev, Ukraine – Sofia, Bulgaria, 2010

ISBN 978-954-16-0048-1

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA  
ITHEA IBS ISC: 19.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligence metasynthesis and knowledge processing in intelligent systems; Formalisms and methods of knowledge representation; Connectionism and neural nets; System analysis and synthesis; Modelling of the complex artificial systems; Image Processing and Computer Vision; Computer virtual reality; Virtual laboratories for computer-aided design; Decision support systems; Information models of knowledge of and for education; Open social info-educational platforms; Web-based educational information systems; Semantic Web Technologies; Mathematical foundations for information modeling of knowledge; Discrete mathematics; Mathematical methods for research of complex systems.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

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

Printed in Ukraine

**Copyright © 2010 All rights reserved**

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. [www.ithea.org](http://www.ithea.org) ; e-mail: [info@foibg.com](mailto:info@foibg.com)

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co., Bulgaria

**ISBN 978-954-16-0048-1**

C/o Jusautor, Sofia, 2010

## PREFACE

ITHEA International Scientific Society (**ITHEA ISS**) is aimed to support growing collaboration between scientists from all over the world. The scope of the books of the ITHEA ISS covers the area of Informatics and Computer Science. ITHEA ISS welcomes scientific papers and books connected with any information theory or its application. ITHEA ISS rules for preparing the manuscripts are compulsory. ITHEA Publishing House is the official publisher of the works of the ITHEA ISS.

Responsibility for papers and books published by ITHEA belongs to authors.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligence metasynthesis and knowledge processing in intelligent systems; Formalisms and methods of knowledge representation; Connectionism and neural nets; System analysis and synthesis; Modelling of the complex artificial systems; Image Processing and Computer Vision; Computer virtual reality; Virtual laboratories for computer-aided design; Decision support systems; Information models of knowledge of and for education; Open social info-educational platforms; Web-based educational information systems; Semantic Web Technologies; Mathematical foundations for information modeling of knowledge; Discrete mathematics; Mathematical methods for research of complex systems.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

The book is recommended for publication by the Scientific Council of the Institute of Information Theories and Applications FOI ITHEA.

Papers in this book are selected from the ITHEA ISS Joint International Events of Informatics "ITA 2010":

<b>CFDM</b>	Second International Conference "Classification, Forecasting, Data Mining"
<b>i-i</b>	International Conference "Information - Interaction – Intellect"
<b>i.Tech</b>	Eight International Conference "Information Research and Applications"
<b>ISK</b>	V-th International Conference "Informatics in the Scientific Knowledge"
<b>MeL</b>	V-th International Conference "Modern (e-) Learning"
<b>KDS</b>	XVI-th International Conference "Knowledge - Dialogue – Solution"
<b>CML</b>	XII-th International Conference "Cognitive Modeling in Linguistics"
<b>INFOS</b>	Thirth International Conference "Intelligent Information and Engineering Systems"
<b>NIT</b>	International Conference "Natural Information Technologies"
<b>GIT</b>	Eight International Workshop on General Information Theory
<b>ISSI</b>	Forth International Summer School on Informatics

ITA 2010 took place in Bulgaria, Croatia, Poland, Spain and Ukraine. It has been organized by

- ITHEA International Scientific Society

in collaboration with:

- ITHEA International Journal "Information Theories and Applications"
- ITHEA International Journal "Information Technologies and Knowledge"
- Institute of Information Theories and Applications FOI ITHEA
- Universidad Politecnica de Madrid (Spain)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Taras Shevchenko National University of Kiev (Ukraine)
- University of Calgary (Canada)
- BenGurion University (Israel)
- University of Hasselt (Belgium)
- Dorodnicyn Computing Centre of the Russian Academy of Sciences (Russia)
- Institute of Linguistics, Russian Academy of Sciences (Russia)
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- New Bulgarian University (Bulgaria)
- The University of Zadar (Croatia)
- Rzeszow University of Technology (Poland)
- Kharkiv National University of Radio Electronics (Ukraine)
- Kazan State University (Russia)
- Alexandru Ioan Cuza University (Romania)
- Moscow State Linguistic University (Russia)
- Astrakhan State University (Russia)

as well as many other scientific organizations. For more information: [www.ithea.org](http://www.ithea.org) .

We express our thanks to all authors, editors and collaborators as well as to the General Sponsor.

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

*Kiev – Sofia  
September 2010*

*K. Markov, V. Velychko, O. Voloshin*

## TABLE OF CONTENTS

<b>Preface</b> .....	<b>3</b>
<b>Table of Contents</b> .....	<b>5</b>
<b>Index of Authors</b> .....	<b>9</b>
<b>PHILOSOPHY AND METHODOLOGY OF INFORMATICS</b>	
<b>Проблема эффективности междисциплинарных исследований</b>	
Александр Кургаев, Александр Палагин .....	11
<b>О шагах системного синтеза (продолжение «Методологического меморандума» – часть II)</b>	
Анатолий Крисиллов, Авенир Уемов .....	18
<b>Модель универсума</b>	
Мержвинский Анатолий .....	31
<b>Ситуационный центр как инструмент для проведения форсайтных исследований</b>	
Виталий Вишневецкий, Сергей Симонов .....	40
<b>Обобщенная оценка организационных систем</b>	
Виталий Вишневецкий, Владимир Калмыков, Олег Малышев .....	46
<b>Моделирование сложных искусственных систем с использованием концепции «искусственная жизнь»</b>	
Богдан Мысник .....	51
<b>О целостности образа: доформальное исследование</b>	
Юрий Валькман .....	58
<b>К интеграции онтологий предметных областей</b>	
Александр Палагин, Андрей Михайлюк, Виталий Величко, Николай Петренко .....	69
<b>Компьютерные онтологии и онтолого-управляемая архитектура информационных систем</b>	
Николай Петренко .....	86
<b>Использование специализированной лексической онтологии для автоматизации формирования онтологии предметной области по естественно-языковым текстам</b>	
Ольга Лесько, Юлия Рогушина .....	93
<b>Критерии принятия решений при выборе редактора онтологий</b>	
Дмитрий Климань, Вера Любченко .....	101
<b>Integration of Financial Domain Knowledge on Base of Semantic Web Technologies</b>	
Anatoly Gladun, Julia Rogushina, Rodrigo Martínez-Béjar, Francisco García-Sánchez and Rafael Valencia-García .....	106
<b>A Comprehensive Approach and User-oriented Policy for Managing Risk and Safety</b>	
Kristian Milenov, Krassimir Markov, Venko Bozhilov, Radko Radkov, Pavel Milenov .....	113
<b>Analogies Between Texts: Mathematical Models and Applications in Computer-assisted Knowledge Testing</b>	
Leonid Leonenko .....	128
<b>On Semantics and Syntax of the BSDT Primary Language</b>	
Petro Gopych .....	135

## NEURAL NETWORKS AND PATTERN RECOGNITION

<b>Сравнительный анализ методов многокритериальной оптимизации архитектуры нейросетевых классификаторов</b>	
Альберт Воронин, Юрий Зиятдинов, Анна Антонюк.....	146
<b>Корректные модели алгоритмов распознавания</b>	
Виктор Краснопрошин, Владимир Образцов .....	154
<b>Метод выделения признаков на основе совместных векторов моментных фазовых распределений и моментов Зернике</b>	
А.Н. Чернодуб .....	162
<b>Обучение рекуррентной нейронной сети методом контролируемого возмущения для управления динамическими объектами</b>	
Д. А. Дзюба, А. Н. Чернодуб .....	170
<b>Исследование модификации FOTSK нечеткой нейронной сети TSK в задачах прогнозирования</b>	
Юлия Проскурня, Богдан Гривко .....	177
<b>Прогнозирование в деятельности предприятия с использованием нейронных сетей</b>	
Галина Сетлак .....	186
<b>Создание вычислительных моделей в области горения конденсированных систем средствами data mining</b>	
Виктор Абриков, Сергей Абриков, Елена Карлович.....	193
<b>Protein Structure Prediction on a Three-dimensional Triangular Lattice</b>	
Leonid Hulianytskyi, Vitalina Rudyk.....	198

## DECISION MAKING

<b>Individually Optimal Principles of Distribution of Greenhouse Gas Emmision Quotas</b>	
Olexij Voloshin, Sergij Maschenko .....	209
<b>Нечеткие обобщения модели распределения затрат</b>	
Алексей Ф. Волошин, Василий А. Лавер.....	215
<b>Нечеткая модель удовлетворительного решения задачи выбора</b>	
Николай Н. Маляр .....	220
<b>О принятии решений при анализе малых возмущений линейных моделей</b>	
Всеволод Богаенко, Владимир Кудин .....	226
<b>Исследование модифицированной системы Мартингейл для трейдинга на валютных рынках</b>	
Юрий Зайченко, Владислав Плаксин .....	232
<b>Прогнозирование риска банкротства банков в условиях неопределенности с применением нечетких нейронных сетей</b>	
Юрий Зайченко, Ольга Войтенко .....	236

<b>Построение и исследования нечеткой модели потребления</b>	
Олег Пошелюжный.....	240
<b>Параллелизация кооперативных модели-ориентированных метаэвристик</b>	
Сергей Сиренко .....	246
<b>Практическое использование виртуальной лаборатории для расчета надежности приборов и резервированных систем</b>	
Александр Палагин, Владимир Романов, Игорь Галелюка, Крассимир Марков, Виталий Величко, Александра Ковырёва, Петер Станчев, Крассимира Иванова, Илия Митов.....	252
<b>Исследование информационных процессов для эффективного воспроизведения дактильного жестового языка</b>	
Юрий Крак, Юрий Кривонос, Богдан Троценко .....	262
<b>A Graph Model Of Shocked Financial System</b>	
Velina Slavova, Lyubomir Mirchev, Hristian Elefteridis .....	272
<b>Business Informatics Study in CEE Counties</b>	
Vasyl Grygorkiv, Ruslan Biloskursky, Andrii Verstiak, Igor Vinnychuk .....	284
 <b>MODERN (e-) LEARNING</b>	
 <b>Современные проблемы образования</b>	
Алексей Ф. Волошин .....	290
<b>Проблемы вхождения Украины в европейское образовательное пространство</b>	
Олег К. Закусило.....	297
<b>Знание-ориентированные системы разработки электронных курсов обучения</b>	
Александр Палагин, Юрий Тихонов, Николай Петренко, Виталий Величко.....	304
<b>Дидактический дизайн системы обучения нового поколения</b>	
Елена Вахтина, Александр Вострухин.....	314
<b>Образовательные порталы для школ – ожидания и реальность</b>	
Диана Богданова .....	323
<b>Элементы адаптивной технологии оценивания знаний в режиме реального времени</b>	
Константин Юрченко, Виталий Снитюк.....	331
<b>Организация процесса индивидуализированного обучения на базе адаптивной системы дистанционного обучения и контроля знаний EduPro</b>	
Павло Федорук .....	335
<b>Образование 2.0: тенденции и технологии</b>	
Любовь Шарабаева, Ольга Захарова .....	341
<b>Мультимедийные образовательные системы: опыт внедрения</b>	
Тамара Клебанова, Роман Яценко, Ольга Захарова .....	347
<b>«Открытые» ответы в компьютерном тестировании знаний</b>	
Леонид Леоненко .....	355

## MATHEMATICAL FOUNDATIONS

<b>Евклидовы пространства: Конструктивные методы описания базовых структур и их использование</b>	
Владимир Донченко .....	362
<b>Мультимножества – альтернатива теоретико-множественной платформы в математических основаниях информационных технологий</b>	
Дмитрий Буй, Юлия Богатырёва .....	377
<b>Структура частично упорядоченного семейства мультимножеств</b>	
Дмитрий Буй, Юлия Богатырёва .....	387
<b>Обобщенное исчисление на домене</b>	
Дмитрий Буй, Ирина Глушко .....	392
<b>Disturbance of Statistical Stability</b>	
Igor Gorban .....	398
<b>Multiple Folding of VLSI Regular Structure via Boolean Satisfiability</b>	
Liudmila Cheremisinova.....	411
<b>Simple Constrained Folding of Programmable Logic Arrays of Special Type</b>	
Liudmila Cheremisinova, Irina Loginova.....	420
<b>Application of Pareto Optimization Approach to Observability Problem in Linear Aerodynamic Models</b>	
Andriy Zavorotnyy, Veda Kasyanyuk.....	429
<b>Cycle Routes Optimization for not Full Graph</b>	
Anatoly Panishev, Anton Levchenko .....	435
<b>Modified Branch and Bound Algorithm for Solving the Hamiltonian Rural Postman Problem</b>	
Andriy Morozov, Anatoliy Panishev .....	442
<b>Comparative Analysis of Statistical Properties of the Hurst Exponent Estimates Obtained by Different Methods</b>	
Ludmila Kirichenko, Tamara Radivilova .....	451
<b>Some Properties of Acyclic Compositional Programs</b>	
Tetiana Parfirova, Vadim Vinnik .....	460



## INDEX OF AUTHORS

### Names in Russian

А. Н. Чернодуб	162, 170	Елена Карлович	193
Авенир Уемов	18	Игорь Галелюка	252
Александр Вострухин	314	Илия Митов	252
Александр Кургаев	11	Ирина Глушко	392
Александр Палагин	11, 69, 252, 304	Константин Юрченко	331
Александра Ковырёва	252	Крассимир Марков	252
Алексей Ф. Волошин	215, 290	Крассимира Иванова	252
Альберт Воронин	146	Леонид Леоненко	355
Анатолий Крисилов	18	Любовь Шарабаева	341
Андрей Михайлюк	69	Мержвинский Анатолий	31
Анна Антонюк	146	Николай Н. Маляр	220
Богдан Гривко	177	Николай Петренко	69, 86, 304
Богдан Мысник	51	Олег К. Закусило	297
Богдан Троценко	262	Олег Малышев	46
Василий А. Лавер	215	Олег Пошелюжный	240
Вера Любченко	101	Ольга Войтенко	236
Виктор Аbruков	193	Ольга Захарова	341, 347
Виктор Краснопрошин	154	Ольга Лесько	93
Виталий Величко	69, 252, 304	Павло Федорук	335
Виталий Вишневский	40, 46	Петер Станчев	252
Виталий Снитюк	331	Роман Яценко	347
Владимир Донченко	362	Сергей Аbruков	193
Владимир Калмыков	46	Сергей Симонов	40
Владимир Кудин	226	Сергей Сиренко	246
Владимир Образцов	154	Тамара Клебанова	347
Владимир Романов	252	Юлия Богатырёва	377, 387
Владислав Плаксин	232	Юлия Проскурня	177
Всеволод Богаенко	226	Юлия Рогущина	93
Галина Сетлак	186	Юрий Валькман	58
Д. А. Дзюба	170	Юрий Зайченко	232, 236
Диана Богданова	323	Юрий Зиатдинов	146
Дмитрий Буй	377, 387, 392	Юрий Крак	262
Дмитрий Климань	101	Юрий Кривонос	262
Елена Вахтина	314	Юрий Тихонов	304

**INDEX OF AUTHORS****Names in English**

Anatoliy Panishev	442	Ludmila Kirichenko	451
Anatoly Gladun	106	Lyubomir Mirchev	272
Anatoly Panishev	435	Olexij Voloshin	209
Andrii Verstiak	284	Pavel Milenov	113
Andriy Morozov	442	Petro Gopych	135
Andriy Zavorotnyy	429	Radko Radkov	113
Anton Levchenko	435	Rafael Valencia-García	106
Francisco García-Sanchez	106	Rodrigo Martínez-Béjar	106
Hristian Elefteridis	272	Ruslan Biloskursky	284
Igor Gorban	398	Sergij Maschenko	209
Igor Vinnychuk	284	Tamara Radivilova	451
Irina Loginova	420	Tetiana Parfirova	460
Julia Rogushina	106	Vadim Vinnik	460
Krassimir Markov	113	Vasyl Grygorkiv	284
Kristian Milenov	113	Veda Kasyanyuk	429
Leonid Hulianytskyi	198	Velina Slavova	272
Leonid Leonenko	128	Venko Bozhilov	113
Liudmila Cheremisinova	420, 411	Vitalina Rudyk	198

---

---

## Philosophy and Methodology of Informatics

---

---

### ПРОБЛЕМА ЭФФЕКТИВНОСТИ МЕЖДИСЦИПЛИНАРНЫХ ИССЛЕДОВАНИЙ

Александр Кургаев, Александр Палагин

**Аннотация:** Сформулирована проблема эффективности междисциплинарных научных исследований; выявлены наиболее существенные признаки научного знания, ограничивающие область оптимума производительности труда ученых; выявлена иерархия проблемных ситуаций на пути роста производительности труда ученых; обоснован путь разрешения этих проблемных ситуаций.

**Ключевые слова:** междисциплинарные научные исследования; ставшее научное знание; производительность труда ученых; оптимизация производительности труда ученых.

**ACM Classification Keywords:** A.0 General Literature; J.4 Social and Behavioral Sciences; M.4 Intelligence Metasynthesis and Knowledge Processing in Intelligent Systems

---

#### Введение

Развитие мирового сообщества постоянно сопровождается множеством проблем, лишь часть которых получает общественное признание, формулируется в явной форме и консолидирует научные, организационные, материальные и финансовые усилия общества на их решение в форме научно-технических программ (НТП).

Накоплен многочисленный, в целом, позитивный опыт формирования и выполнения НТП решения многих проблем [Кун, 1977]. Однако наравне со значительным позитивным опытом формирования и выполнения НТП имеют место определенные недостатки как *организации*, так и, собственно, самого *проведения* научных исследований, особую остроту приобретающие при решении проблем практики развития общества, необходимо требующих междисциплинарных исследований, усложняющих труд ученых до почти непреодолимого барьера в связи со сверхвысокой сложностью их специфического содержания. Их следствием является потеря части народнохозяйственного эффекта от выполнения каждой из НТП и всего их множества.

**Задачей** статьи является выявление проблемных ситуаций на пути прогресса междисциплинарных исследований и наиболее существенных признаков информационных технологий, адекватных естественным процессам решения проблем.

---

#### 1. Характеристика сущности междисциплинарных исследований

Теоретическое знание функционирует и развивается как сложная система внутри дисциплинарных и междисциплинарных взаимодействий, целенаправленных на решение актуальных проблем эволюции науки и практики общества. Общепризнано, что эволюционное развитие науки время от времени прерывается революционными изменениями схемы (*парадигмы*) деятельности познания, воплощающей прогрессивную систему идеалов и норм исследования [Кун, 1977].

Всякий познавательный процесс имеет двунаправленную природу – от менее к более содержательному и от менее к более общему знанию. Стремление к единству знания при максимальном охвате реальности обусловлено единством мира, к отображению которого в науке это знание направлено, а рост

содержательности – непрерывным проникновением науки во все более глубокую сущность действительности, определяя все более адекватную ее модель [Поппер, 1983].

В самом общем виде внутри дисциплинарный генезис знания включает две составляющие: от частных теорий, решающих специфичные проблемы части объектов дисциплины, – к фундаментальной теории, фиксирующей самые общие знания относительно всей области дисциплины, и от фундаментальной теории – к частным теориям, углубляющим ее знания.

Всякое междисциплинарное исследование включает: выявление новых отношений между терминами исходных дисциплин, установление новой системы законов, связывающей их, и синтез прагматики решения новых, все более сложных задач. При этом знания исходных дисциплин могут остаться неизменными (простейший, *линейный* случай междисциплинарного взаимодействия), включенными целиком (или выборочно) в новую иерархическую структуру, или подвергнуться модификации, развитию за счет процессов обмена парадигмальными установками, понятиями и методами между разными науками (*нелинейное* междисциплинарное взаимодействие), диктуемыми междисциплинарным характером проблемы. Но во всех случаях происходит углубление научных знаний и расширение, уточнение области объектов действительности, исследуемых исходными дисциплинами [Степин, 2003].

Нелинейное междисциплинарное взаимодействие в сочетании с проблемной ориентацией уже стало общепризнанной нормой глобальной эволюции современной науки ([Степин, 2003], [Кургаев, 2008]).

В результате этого процесса, де-факто, формируется новая дисциплина и бесконечно повторяется итерационная последовательность событий (от одних дисциплин через междисциплинарное исследование – к новым дисциплинам с новым предметом исследования) в направлении приближения к целостной общенаучной картине мира.

Реальное положение дел в науке характеризуется тем, что проблемные ситуации, возникающие на пути бесконечного процесса познания, непрерывно усложняются, включают все большую область явлений действительности и все большее их число для своего разрешения требует проведения междисциплинарных исследований ([Степин, 2003], [Кургаев, 2008], [Палагин, Кургаев, 2009]). [

На основании высказанного, наиболее существенными признаками междисциплинарных исследований можно считать следующие:

- *активное владение* широким кругом знаний;
- *унификация формы* представления знаний разных научных дисциплин применительно к решению актуальной проблемы;
- *недетерминизм* творческого процесса содержательного соединения недостаточно конструктивных научных знаний, накопленных разными дисциплинами, в такое концептуально единое новое знание, создание которого и является целью соответствующей НТП.

Сложившееся состояние дел не может существовать долго, поскольку непрерывно усложняет научную работу, что ведет к постоянному снижению эффективности труда ученых, вследствие чего снижается значимость новых знаний и, в конечном итоге, уменьшается вклад науки в решение неотложных проблем эволюции общества, угрожая его полной потерей.

На основании изложенного, сущность проблемы эффективности исследований, в конечном итоге, сводится к следующей **проблеме 0-го** уровня:

*Достижение общественно значимого роста производительности труда ученых за счет усовершенствования информационных технологий, целенаправленного на преодоление преград на пути прогресса науки.*

---

## 2. Необходимость решения проблемы в форме НТП

---

Сформулированная **проблема 0-го** уровня – фундаментальна, ее решение зависит от целого комплекса условий (научных, организационных, материально-технических, финансовых и кадровых), среди которых

определяющим является творчество специфичных знаний многих научных дисциплин. Оставаясь основным предметом методологии науки, процессы познания и их компоненты исследуются при решении своих специфичных проблем в кибернетике и математике, теории сложных систем, лингвистике, логике и во всем спектре компьютерных наук: от вычислительной техники до систем обработки знаний и искусственного интеллекта.

**2.1 Непродуктивная составляющая труда ученых.** В самой абстрактной форме научная деятельность осуществляется согласно итерационной структуре (рис. 1).

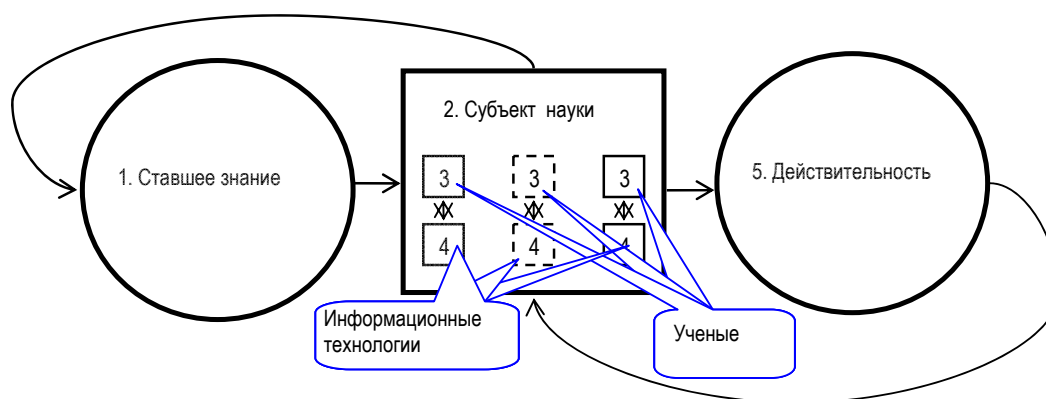


Рис.1. Итерационная структура научной деятельности

Из этого следует непосредственная зависимость производительности труда ученых от трудозатрат на *поиск, выявление, осознание* актуально необходимого фрагмента ставшего знания и *оперирования* им в ходе исследований. В свою очередь, объем всех этих составляющих трудозатрат определяется совершенством формы представления ставшего знания, циркулирующего в обществе, т.е. уровнем его наиболее существенных признаков:

- *известности* (диапазон: от общеизвестных, узкопрофессиональных до неизвестных знаний);
- *явности* (диапазон: от явной до неявной формы);
- *унифицированности* (измеряется разнообразием форм);
- *конструктивности* (диапазон: от пригодности до непригодности для непосредственного использования при решении проблем).

Очевидно, глобальный минимум трудозатрат или максимум производительности профессионального труда ученых существует лишь на очень небольшом дискретном подмножестве (относительно совокупного ставшего знания всего пространства науки) знания, – представленном в единой унифицированной форме, явном, конструктивном и общеизвестном относительно всей науки (или общеизвестном относительно дисциплин, к которым принадлежит исследуемая проблема). Форму знаний, отвечающую этим условиям, назовем *канонической* (рис. 2).

Обычно, *оптимальные условия отсутствуют* по многим причинам и, преимущественно, имеет место некоторое сочетание негативных признаков знания: *не явность, не известность* (общая или относительно некоторой дисциплины), *не конструктивность* и *не унифицированность*. Во всех этих случаях, вместо непосредственного исследования текущей проблемы ученые тратят усилия на *поиск* (в условиях полного или частичного отсутствия гарантий существования), *осознание* и *трансформацию* ставшего знания из неявного в явное, из неконструктивного в конструктивное и из разных форм к единой применительно к условиям текущей проблемы.

Все трудозатраты на *поиск*, *осознание* и *трансформацию* ставшего знания, т.е. на его многократную вторичную обработку, являются непродуктивными относительно решения каждой из множества текущих проблем; их объем достигает львиной доли совокупных затрат труда на решение отдельных проблем, а в сумме могут значительно превышать трудозатраты на первичное творчество ставшего знания. Более того, эта часть труда ученых не только общественно бесполезна, но даже наносит вред, поскольку в каждом конкретном исследовании выполняется *ad hoc* (т.е. относительно специфических условий текущей проблемы), непрерывно увеличивая энтропию ставшего знания в общем пространстве науки.



Рис. 2. Область максимума производительности профессионального труда ученых (общая площадь пяти колец) – лишь малая часть ставшего знания общего пространства науки

**2.2 Альтернативные пути решения проблемы.** На всем множестве возможных путей решения **проблемы 0-го уровня** есть две альтернативы: эволюционный – постепенным накоплением позитивных эффектов и революционный – за счет парадигмальных нововведений в развитие информационных технологий обработки и распространения научных знаний в обществе.

Эволюционному пути присущ ряд недостатков, в частности:

- поскольку процесс познания *недетерминирован*, а *пригодность* результатов науки (в современной форме) *к контролю* слишком *низка*, каждое исследование вносит в общую копилку две части знания – ту, что уменьшает энтропию информированности о действительности, и ту, что ее увеличивает. При этом состояние развития цивилизации не добавляет уверенности в уменьшении общей энтропии знаний при использовании наличных информационных технологий. Даже, если уменьшение и имеет место, то вполне очевидно, что его *темпы не соответствуют требованиям времени*;
- вследствие информационного "взрыва" и непрерывного расширения научного пространства, *объем ставшего знания уже сейчас достиг критической черты* способности ученых на получение актуально необходимой части знания, усвоение и активное его использование;

- принципиально *неприемлема компенсация* непродуктивной составляющей труда ученых за счет *экстенсивного развития* науки;
- допустимо утверждать, что современное *состояние развития* науки близко к *кризисному*.

Тем самым налицо проблемная ситуация, не разрешимая при текущем положении дел. Поскольку острота проблемной ситуации побуждает к решительным действиям для ее разрешения, указанные недостатки и трудности эволюционного развития событий вынуждают сосредоточить усилия на разработке и реализации альтернативного пути, тем более что по общему убеждению *времени на промедление у науки нет!*

В целом, путь парадигмальных нововведений вполне естественный, подтвержден фактами развития как отдельных научных дисциплин, так и всей науки, и многократно себя оправдал.

**2.3 Основная идея решения проблемы.** Предлагается кардинально изменить методы добычи, представления и использования научного знания и, как следствие, – отношение к знанию со стороны общества, самих ученых и, в конечном итоге, к науке в целом [Палагин, Кургаев, 2009].

В современном обществе общепринято убеждение, что *научное знание* (независимо от его профиля, уровня и места, времени создания и субъекта-творца) является *разновидностью сырья*, использование которого *невозможно* без затрат дополнительных ресурсов на его переработку в процессе творчества нового знания и/или создания более или менее полезной материальной или информационной продукции. Оснований для него вполне достаточно, о чем свидетельствует практика мировой науки, а все ученые смиренно принимают его как неизбежное зло или, наоборот, как общественное разрешение на несовершенство результатов личного труда.

Очевидно, такое состояние дел не удовлетворяет никого. Во взаимодействии общества с наукой каждый заинтересован получить наибольший вклад в свое развитие: общество от науки, а наука от общества. Выход из этой проблемной ситуации есть лишь один – пойти навстречу друг другу, наполнив конструктивным содержанием и приняв к совместной реализации общеизвестный тезис: "*Наука – непосредственная производительная сила прогресса общества!*". В основу новой парадигмы предлагается положить фундаментальное утверждение.

**Утверждение.** *Ставшее знание в канонической форме – самая совершенная разновидность основного капитала общества, обеспечивающая развитие науки и прогресс цивилизации.*

Основания для него есть и сейчас, поскольку его составляющим присущи общие признаки (пригодность для тиражирования, модификации, развития, улучшения безопасности и комфорта жизни и труда, роста темпов прогресса общества, образования прибыли и др.). Относительно же качества перечисленных признаков, безусловно, что у знаний они существенно лучше, чем у материальных разновидностей основного капитала, причем только научное знание в состоянии определить и непрерывно корректировать развитие цивилизации в направлении ее гармонии с природой. Несмотря на это, очевидно, что относительно современности предложенное утверждение ложно, поскольку у научного знания еще отсутствуют такие наиболее существенные признаки конечного товара и условия для его существования в обществе, как:

- пригодность для *непосредственного использования*;
- *измеримость* показателей количества и качества;
- рыночная *цена и стоимость*;
- *действенные нормы прав и обязанностей* производителя и потребителей.

Суть решения **проблемы 0**-го уровня состоит в воплощении необходимых и достаточных условий для действенного функционирования ставшего знания как товара и стимулирования этого процесса. Главным

средством достижения этих условий должно стать решение множества научных проблем относительно создания системы новых компьютерных информационных технологий, адекватных естественным процессам решения проблем в науке.

**2.4 Иерархия проблемных ситуаций.** На пути решения **проблемы 0-го уровня** возникает группа **проблем 1-го уровня**, о наличии которых свидетельствуют проблемные ситуации 1-го уровня, подтверждаемые фактами современного состояния науки:

- *неадекватность* компьютерной обработки знаний естественному процессу научных исследований;
- *отличие формы* научного знания, используемого для конструктивного решения проблем от той формы, что циркулирует в инфраструктуре распространения знания в обществе.

Каждая из указанных проблемных ситуаций 1-го уровня, в свою очередь является комплексной и требует проведения исследований разных научных дисциплин для решения **проблем 2-го уровня**, о наличии которых свидетельствуют, в частности, проблемные ситуации:

- *недостаточная* конструктивность научных знаний, их *непригодность* для непосредственного использования для решения текущих и новых проблем и внедрения в практику;
- *разнообразие* существующих форм ставшего знания (т.е. структур и языков теорий), присущих разным научным дисциплинам, в сравнении с *унифицированной* формой концептуально единого нового знания, для творчества которого и формируется соответствующая НТП;
- *сложность, разнообразие и неопределенность* естественных творческих процессов постановки и решения научных проблем, их разновидностей и компонент;
- *неадекватность* компьютерных информационных технологий, моделей и языков представления знаний естественным творческим процессам постановки и решения научных проблем;
- *неадекватность* действующей Internet-технологии потребностям ученых относительно точности и релевантности поиска актуально необходимого научного знания;
- *производство* нового научного знания осуществляется интенсивным трудом творческих личностей и достается *дорогой ценой*, а *продается* излишне *дешево*, почти по цене носителя, истраченного на фиксацию этого знания.

---

### 3. Ожидаемые результаты

---

В результате выполнения исследований, направленных на разрешение перечисленных проблемных ситуаций, должны быть определены концептуально единое пространство ставшего знания в каноничной форме и совокупность адекватных взаимно согласованных информационных технологий эффективной поддержки всех составляющих профессионального труда ученых по конструктивному решению произвольных научных проблем и непосредственному внедрению нового научного знания в практику.

Анализируя суть причин и следствий парадигмальных нововведений в развитие компьютерных систем, можно удостовериться на многих фактах, что вся эволюция компьютерной науки связана с получением именно этих результатов. Если при зарождении компьютеров основное внимание было сосредоточено на вопросах осуществимости конструктивных вычислений и достижении рекордных значений количественных показателей, то со временем на передний план вышли вопросы *адекватности* информационно-аналитической поддержки всех составляющих естественного процесса решения научно-практических проблем. В настоящее время основное внимание сосредоточено на реализациях систем обработки знаний, представленных разными формальными моделями с использованием современных математических методов и архитектурных решений, в частности, с аппаратной поддержкой этих методов.

За счет придания ставшему знанию сочетания наиболее существенных креативных признаков, в условиях всесторонней действенной их поддержки средствами адекватных информационных технологий и



правовыми нормами, научные знания приобретут упомянутые выше, но отсутствующие сейчас, признаки конечного товара:

- пригодность для *непосредственного использования*;
- *измеримость* показателей количества и качества;
- рыночную *цену и стоимость*.

Эффективность разрешения приведенных выше проблемных ситуаций определяется суммой эффектов (в целом разного качества) от приобретения этих и других признаков.

---

## Заключение

### Обосновано:

- область максимума производительности профессионального труда ученых ограничена ставшим знанием в канонической форме, т.е. явным, конструктивным и представленным в унифицированной форме;
- междисциплинарный характер сформулированной **проблемы 0**-го уровня;
- целесообразны парадигмальные нововведения в развитие информационных технологий обработки и распространения научных знаний в обществе.

**Предложено** принять к воплощению новое отношение к научному знанию (как к *самой совершенной разновидности основного капитала общества*) за счет придания ему отсутствующих в настоящее время наиболее существенных признаков конечного товара и создав наиболее благоприятные условия для его существования и развития в научной и общественной практике.

---

## Библиография

- [Кун, 1977] Кун Т. Структура научных революций / Пер. с англ. – М.: Прогресс, 1977. – 300 с.
- [Поппер, 1983] Поппер К. Логика и рост научного знания. Избранные работы / Пер. с англ. – М.: Прогресс, 1983. – 496с.
- [Степин, 2003] Степин В.С. Теоретическое знание. Структура, историческая эволюция. – М.: Прогресс-Традиция, 2003. – 744 с.
- [Кургаев, 2008] Кургаев А.Ф. Проблемная ориентация архитектуры компьютерных систем. – Киев: Сталь, 2008. – 540с.
- [Палагин, Кургаев, 2009] Палагин О.В., Кургаев О.П. Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки // Вісник Національної академії наук України. – 2009. – № 3. – С. 14-25.

---

## Информация об авторах



**Александр Кургаев** – доктор технических наук; ведущий научный сотрудник Института кибернетики имени В.М. Глушкова НАН Украины; проспект Глушкова, 40, Киев 187, Украина; e-mail: [afkurgaev@ukr.net](mailto:afkurgaev@ukr.net)

Основные области научных исследований: информационные машины и системы обработки знаний, представленных в форме научных теорий



**Александр Палагин** – академик НАН Украины; заместитель директора Института кибернетики имени В.М. Глушкова НАН Украины; проспект Глушкова, 40, Киев 187, Украина; e-mail: [palagin\\_a@ukr.net](mailto:palagin_a@ukr.net)

Основные области научных исследований: системная интеграция трансдисциплинарных научных знаний, онтологический инжиниринг

## О ШАГАХ СИСТЕМНОГО СИНТЕЗА (ПРОДОЛЖЕНИЕ «МЕТОДОЛОГИЧЕСКОГО МЕМОРАНДУМА» – ЧАСТЬ II)

**Анатолий Крисилов, Авенир Уемов**

**Аннотация:** Рассмотрены методологические вопросы и последовательность построения процедур синтеза, роль самоопределения и системных дескрипторов. Описаны разновидности синтеза: реистический, атрибутивный, реляционный, а также некоторые цели и задачи синтеза.

**Ключевые слова:** системный анализ и синтез, самоопределение, системные дескрипторы, виды синтеза: реистический, атрибутивный, реляционный; задачи синтеза.

**Summary:** The work is devoted to the system study of the complicated and weak-formalised objects, to its synthesis especially. The methodological approach to describing of system descriptors, of synthesis' steps and such kinds of synthesis as reistic, attributive and relation ones is used. A good deal of attention is paid to the synthesis' goals and tasks.

**Keywords:** system analysis and synthesis, system descriptors, kinds of synthesis: reistic, attributive and relation ones.

**ACM Classification Keywords:** H.1.1 – Systems and Information Theory – General Systems Theory.

*Там на неведомых дорожках*

*Следы невиданных зверей...*

*А.Пушкин*

*Среди невзятых рубежей...*

*В. Высоцкий*

---

### 1. Вводные замечания

---

На киевской (осенней) сессии KDS-2009 была представлена первая часть настоящей работы, названная «Кратким методологическим меморандумом» [1, 2]. Она была посвящена изложению основных идей системного анализа с точки зрения системологии: в ней говорилось о приемах и методах системного исследования сложных и слабоформализованных объектов, имеющих, в основном, качественную природу. Были рассмотрены основные этапы системного анализа как эффективного инструмента для описания, структурирования, моделирования именно таких объектов и процессов.

В настоящей работе, являющейся идеологически продолжением предыдущей, содержится уточнение и развитие некоторых положений, описанных в первой части «Меморандума». Это относится, в частности, к задаче **самоопределения** (описаны вопросы/приемы оценки значимости рассматриваемой ситуации или системы до того, как проводить ее детальный анализ), к задаче описания сложных объектов при помощи системных дескрипторов, к задачам **моделирования**, к описанию некоторых **прикладных и методологических задач**, на решение которых может быть ориентирован системный синтез.

В работе [1] подчеркивалось, что по существу системный анализ должен представлять собой совокупность методов исследования (описания) таких свойств и отношений в системе, которые характеризуют ее именно как целенаправленную, совокупность средств и направлений (!) **изучения связей в этой системе как взаимоотношений между ее целями и способами их реализации.**

Системный синтез, опирающийся на результаты анализа и органически вытекающий из него (будучи формально разложенным на составляющие и описанным!), – представляет, таким образом, инструмент для возможно более полного описания и построения сложных систем (в частности, в «интеллектуальной» компьютерной среде, но не только), для конструирования из баз данных – баз знаний, для обнаружения в известных системах новых сторон, качеств и связей, для синтезирования, наконец, новых знаний.

Этим вопросам в настоящее время посвящено значительное число работ, например, по мультиагентным системам и сложным средам, по разработке различных онтологий, их редукции и расширению и т. д. Многие из этих исследований носят конкретный прагматический характер, их результаты успешно применяются. Следует сказать, однако, что почти все они построены в ключе определенных «компьютероморфных» представлений: для компьютерной реализации, работы с компьютерной моделью, с учетом вычислений на компьютере. В то же время хотелось бы посмотреть на задачи описания и синтезирования знаний с более общих, с системных позиций, больше исходя из «человеко-морфных» и собственно «системно-морфных» представлений. Определенный объем работы в этом направлении разными авторами уже проделан, однако, надо думать, что пока мы находимся в начале пути.

Практически для большинства управленческих и организационных задач (и, что очень важно, для задач ноосферных) следует рассматривать систему вместе с ее окружением, надсистемой. В этом случае под системой понимается функциональный объект, свойства которого обусловлены функцией этого объекта, **сводящейся, в конечном счете, к поддержанию определенных свойств надсистемы** [3]. Здесь четко возникает представление об иерархичности рассматриваемой связи, но не столько как об управленческой иерархии, а как об иерархии структурной, отражающей, например, иерархию смыслов, степени важности задач и т. д. И тогда на один из верхних уровней (для исследователя или конструктора) выходит задача анализа функций объекта. Эта группа обстоятельств (**учет охватываемой системы, важность анализа функций, понимание взаимной вложенности и координации смыслов**) – является чрезвычайно важной при решении задач системного синтеза.

Разработанный для проектных задач СВОН-анализ, который нацелен на учет позитивных и негативных сторон (например, проекта), на описание лимитирующих и продвигающих факторов, – успешно применяется в конкретных прикладных работах, где объект поддается такому описанию, многое известно, вопросов о целях, границах, связях не возникает, и где оценки вообще черно-белые. Но если ситуация сложная, барьеры внешние или трудности внутренние не очевидны, влияющие факторы многомерны, – названный инструментальный эффективного результата не даст, следует применять системный анализ.

Завершая эти краткие вводные замечания, отметим, что в предлагаемых работах по системному анализу и синтезу, при изложении их операциональных шагов, сделана попытка хотя бы конспективно описать определенный (системный, системологический) инструментальный, применение которого в прикладных областях сможет подсказать разработчику новые связи, новые аспекты поиска, возможность составить более полное описание исследуемого объекта или придать конструируемой системе новые качества.

---

## 2. Некоторые методологические соображения и этапы системного анализа

---

Выскажем несколько общих (методологических) соображений об операциях анализа и синтеза, о визуальной модели построения целостного знания, о роли самой методологии. Начнем с последнего.

В работе [1] говорилось о существенной разнице между методикой и методологией: если методика, методическое руководство отвечает на вопрос: **«Как нечто надо делать?»**, то методология отвечает на вопрос: **«Что в этом нечто надо делать?»** Это означает, что методология определяет, какие задачи следует решить, чтобы представить рассматриваемый объект как целостное, единое образование, с его окружением (то есть что представляет собой мир, в котором он существует или будет существовать), с его структурой (то есть совокупностью его элементов и связей между ними), с его свойствами, отношениями между ними, свойствами этих отношений и т. д.

По сути, речь идет о формировании системных знаний об исследуемом объекте – в данной предметной области. И системологический инструментальный является, таким образом, средством для формирования

таких знаний; отметим – data mining и knowledge discovery одновременно. Можно сказать, таким образом, что методология (в данном случае – системология) выступает **одним из мощных средств организации ноосферы как сферы знания.**

Картину формирования целостных знаний об объекте исследования/конструирования, способ получения системного представления о нем, – опишем с помощью определенной геометрической аналогии [4, 5].

Представим себе некоторую плоскость (см. рис. 1), в которой располагается содержательная информация из определенных наук; эту плоскость назовем онтологической. Там лежат предметные представления, например, о географии (где и какие находятся континенты и океаны, где какой климат, где и какая обитает биота и т. д.), об экономике (уклады, предприятия, транспорт, рынки,..), о социальной истории (хронология событий, императоры и военачальники, становление и падение цивилизаций,..), о биологии, о сопротивлении материалов и так далее.

Перпендикулярно к этой, скажем, горизонтальной плоскости существует другая плоскость, вертикальная. В ней находятся совсем другие понятия и категории: принципы, закономерности, механизмы, критерии, понятие о структуре и инфраструктуре и т. д., – это методологическая плоскость.

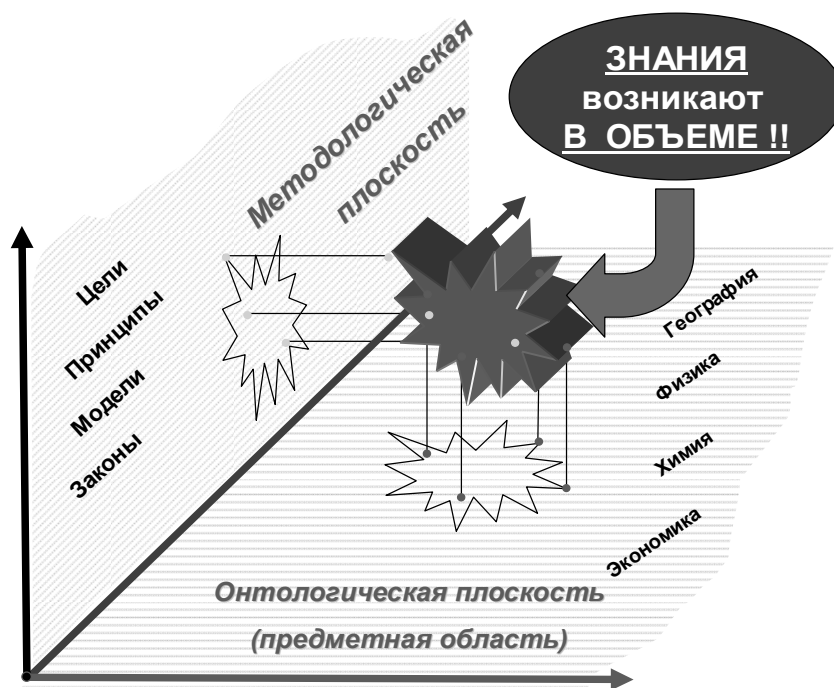


Рис. 1. Модель формирования знаний

Так вот, настоящее **знание**, целостное и системное представление об объекте (процессе, ситуации, явлении), их более полное понимание – **формируется в объеме**, образованном двумя указанными плоскостями в данном фазовом пространстве. В этот объем (скажем, в гиперсферу, репрезентирующую объект данного исследования) из онтологической плоскости берутся конкретные предметные данные об этом объекте (например, имя предприятия, его персонал, номенклатура продукции, прибыль,..), а из методологической плоскости – понятия о его целях, функциях, о характере его структуры, о степени удовлетворения им запросов охватывающей системы и т. д. Из этих организованных определенным способом данных и формируется целостное представление об объекте, системное знание о нем.

Несколько слов об анализе и синтезе. Эти две связанные группы операций в рассматриваемых задачах представляют собой определенное диалектическое единство: дополняют друг друга, переходят друг в друга, могут являться частью друг друга, порой между ними оказываются весьма размытые границы. При

этом, разумеется, в целях определенности в целом ряде случаев и задач устанавливают вполне четкое различие между анализом и синтезом. Это может быть показано на примере трех типов задач, рассматриваемых в радиотехнике или вообще в области связи.

Если есть некоторое устройство  $A$  (напр., усилитель, канал связи, электронный преобразователь,...), есть входной и выходной сигналы  $x(t)$  и  $y(t)$  соответственно, то можно сформулировать три следующие задачи:

1. Известны  $x(t)$  и  $A$  (например, величина сигнала на входе усилителя и коэффициент усиления); задача состоит в том, чтобы рассчитать  $y(t)$  – сигнал на выходе усилителя.
2. Известны  $y(t)$  и  $A$  (например, наблюдаемый сигнал на выходе канала связи и величина помех в канале); задача: восстановить  $x(t)$ , – то есть найти, каким был сигнал на входе канала.
3. Известны  $x(t)$  и  $y(t)$ , т. е., входной и выходной сигналы; задача состоит в том, чтобы построить устройство  $A$ , обеспечивающее требуемое преобразование  $x(t)$  в  $y(t)$ .

Первые две задачи называют задачами анализа, третью – задачей синтеза.

Изложенное показывает, что уже для такого относительно несложного случая разделение операций на анализ и синтез оказывается делом, вообще говоря, конвенционалистским, то есть таким, в котором исследователи должны договориться, что к чему они относят, где пролагают границу.

В рамках наших постановок, как это принято, под анализом понимают расчленение (как правило, мысленное) некоторого объекта на его составные части, выявление связей между ними и т. д.; другое значение понятия «анализ» – оно является прямым синонимом слова «исследование». Синтез же – это соединение, конструирование, объединение (мысленное или реальное) разрозненных элементов в некое единое целое, с требуемыми характеристиками, функциями и т. д.

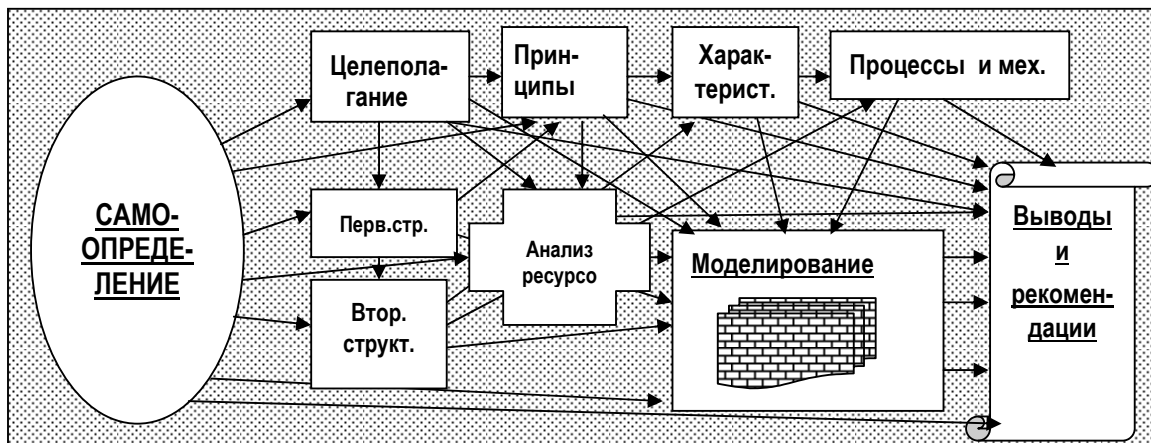


Рис. 2. Основные этапы системного анализа

Прежде, чем заняться элементами синтеза, перечислим кратко основные этапы системного анализа, рассмотренные, в частности, в «Кратком методологическом меморандуме». В него входили следующие этапы (см. рис. 2):

1. Самоопределение.
2. Целеполагание; постановка задачи.
3. Первичная структуризация.
4. Декомпозиция или вторичная структуризация.
5. Выявление основных принципов и закономерностей в данной системе.

6. Основные качественные и количественные характеристики данной системы.

7. Анализ основных процессов в данной системе, имеющих в ней механизмов.

8. Анализ ресурсов.

9. Моделирование.

10. Анализ результатов; формулировка новых целей.

В кратком виде содержание этих этапов было изложено в [1, 2], некоторые дополнения будут сделаны ниже, в том числе, – в контексте задач синтеза.

Рассмотрим теперь, какие этапы, какие процедуры входят в системный синтез.

---

### 3. О шагах системного синтеза

---

Не должно возникать сомнения, что многие позиции из приведенного выше перечня будут входить в число желательных (и обязательных) этапов системного синтеза. В первую очередь, это относится к таким этапам, как Целеполагание и постановка задачи, Самоопределение, Первичная структуризация, Выявление и описание основных принципов и закономерностей, желательных в реализуемой системе... Но, разумеется, во-первых, эти этапы носят несколько иной, нежели при анализе, характер, во-вторых, при синтезе следует выполнить ряд специфических этапов, которые в задачах анализа могут рассматриваться как факультативные, а здесь приобретают законченный конструктивный смысл.

Одним из таких этапов, органически включаемым после целеполагания и являющимся продолжением постановки задачи, является этап формирования (формулирования, задания, определения) совокупности **требований**, предъявляемых к синтезируемому объекту. В этот этап также должна входить работа по определению признаков и критериев, которые дадут возможность оценить степень приближения к целевым состояниям синтезируемой системы.

Перед этим должна быть выполнена другая важная работа – провести **анализ надсистемы**, в которую будет вписан данный синтезируемый объект, и чей запрос он будет своими внешними функциями удовлетворять. Следует полагать, что это может быть один из необходимых элементов самоопределения. Что касается целеполагания, то, как говорилось ранее ([1, 2]), здесь разработчику нужно определить (сформулировать), по крайней мере, три цели (или три группы целей): цели надсистемы, цели синтезируемого объекта и цели собственные; понятно, что это очень емкая и важная работа.

После формирования требований целесообразно выполнить **анализ ресурсов**: чем располагает разработчик, каковы его проектные, инструментальные, модельные, временные и другие возможности.

Далее, в соответствии с требованиями предстоит провести **выбор и реализацию** одного или нескольких **видов синтеза**, предлагаемых параметрическим вариантом общей теории систем: реалистический (предметный, «вещный»), атрибутивный, реляционный; эти разновидности кратко будут рассмотрены ниже. Проведенная к этому моменту работа даст возможность рассмотреть такие задачи, как реализацию основных рабочих **принципов и закономерностей** в системе, **агрегирование** и построение **обобщенных оценок** и характеристик, классификацию и, что очень важно, **таксономию** (выявление или формирование новых классов, номинирование, «именование» таксонов – классификационных единиц). Закрывающими шагами в предлагаемой схеме системного синтеза станут **моделирование, интерпретация и анализ/оценка их результатов**. Некоторые связи между перечисленными шагами показаны на рис. 3.

В более или менее полном виде целесообразный набор шагов, этапов системного синтеза может выглядеть следующим образом:

1. Самоопределение.
2. Целеполагание; постановка задачи.
3. Формирование требований к синтезируемому объекту, системе.
4. Анализ ресурсов.

5. Первичная структуризация (границы системы, входные и выходные параметры).
6. Определение основных принципов и закономерностей для данной системы.
7. Работа с системными дескрипторами (концепт, структура, субстрат).
8. Выбор и реализация видов синтеза (реистический, реляционный, атрибутивный синтез).
9. Описание и встраивание необходимых механизмов.
10. Агрегирование, интегральные оценки.
11. Классификация; таксономия.
12. Моделирование.
13. Интерпретация, учет контекста.
14. Анализ результатов моделирования и интерпретации. Новые цели.

Выбираемый и реализуемый набор действий, «глубина погружения» на каждом шаге – в значительной степени определяются решаемой задачей, однако общая схема представляется именно такой.

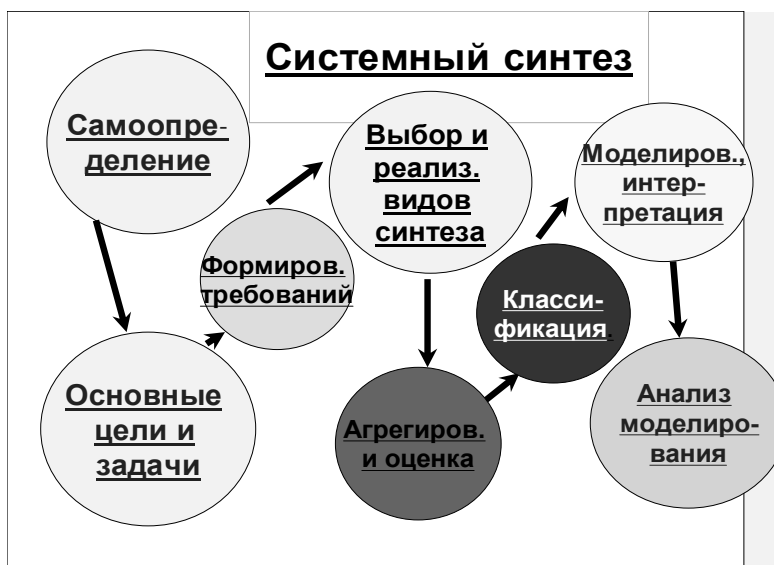


Рис. 3. Некоторые шаги системного синтеза

#### 4. О самоопределении

Самоопределение является чрезвычайно важным и одним из наиболее сложных этапов любой работы и системологической – в особенности. Здесь надо разобраться – кто мы есть по существу, на самом деле? где мы находимся вообще и в данное время? во что мы погружены? (что вокруг нас?) с чем имеем дело?

У Руссо есть такое высказывание: «Истиной сознания является самосознание, но чтобы достичь этой истины, сознание должно охватить то, что лежит вне его», и, добавим, оперировать этим. Помните, как Мюнхгаузен, современник Руссо, вытягивал себя из болота? То, что должен проделать исследователь на этом этапе, очень похоже на действия знаменитого барона, по крайней мере – по сложности. Мысль Ж.-Ж. Руссо хорошо объясняет эту сложность. Вторая теорема Геделя о неполноте говорит: в рамках данной системы можно сформулировать утверждение, истинность или ложность которого средствами самой этой системы проверена быть не может, – надо выходить за ее пределы. Очень любопытное соображение, имеющее несколько другое содержание, но очень близкий смысл, можно найти у Плотина, одного из

первых античных философов-диалектиков: процесс становления вещи – это постоянное восприятие того, что не принадлежит ей, что находится за ее границами.

Каких качеств требует выполнение самоопределения? *Объективность, широта охвата, умение непредвзято видеть себя со стороны, моделировать чужое мышление, проявить последовательность, логичность, самокритичность, «самоответственность»*, – вот примерный букет необходимых качеств.

Правильная и полная реализация этого этапа в очень большой степени определяет успех всей работы.

Среди задач, которые следует решить на этапе самоопределения, выделим такую, как *оценка значимости синтезируемого объекта или решаемой задачи*. Понятно, что, поскольку мы находимся еще на самом начальном этапе, и об исследуемом объекте может быть известно еще очень немного, – такую оценку значимости, вероятно, придется пока делать интуитивно. Формализацию сходного оценочного процесса применительно к работе мыслительной системы описывал В. В. Налимов [6]. Можно сказать, что существует «определенное семантическое поле  $S$ , заданное (как функции – А. К. и А. У.) на всем множестве рассматриваемых смыслов  $T$ . В процессе решения некоторой конкретной задачи спонтанно возникает фильтр (у того, кто ее решает – авт.), мультипликативно взаимодействующий с функцией  $S(T)$ ». Этот фильтр В. Налимова может пониматься как некая призма восприятия (см. рис. 4). На этой основе и создается *система частных оценок* для решения *новой задачи* или старой, но уже в новых условиях [4].



Рис. 4. Восприятие наблюдаемых ситуаций разными «позиционерами».

Несмотря на сложности, и на этом этапе оказывается возможным, опираясь на некоторые подсказки, получить более или менее справедливую оценку значимости. Например, целесообразно иметь в виду, что эта значимость может быть рассмотрена с позиций разных действующих лиц или систем. Скажем, строительство глубоководного выхода «Дунай – Черное море» очень по-разному должно оцениваться отраслевыми структурами и региональными властями; участие или неучастие данной городской команды в чемпионате страны будет иметь несколько различную значимость для болельщиков этой команды и для спортивной общественности в целом. То есть следует попытаться ответить на вопрос «Значимость для



кого?» и, например, хотя бы по трехбалльной шкале (высокая, средняя, низкая) оценить ее для разных подразумеваемых позиционеров.

Далее, определить, на какой временной горизонт рассчитано функционирование данной системы, насколько она актуальна. Очень важно понять, какие потребности (какие уровни в пирамиде Маслоу, в пирамиде социальных потребностей) она удовлетворяет, – физиологические, потребность безопасности, признания, самореализации; для этого целесообразно продумать шкалы этих потребностей для коллектива, для социума. В результате таких (непредвзятых) рассуждений может быть получена некая предварительная и обобщенная оценка значимости, с ее последующей целенаправленной коррекцией.

---

## 5. О роли системных дескрипторов

---

В предыдущей работе [1] применительно к такому этапу, как первичная структуризация, упоминались системные дескрипторы. Проведем это рассмотрение несколько более полно и более корректно – в системном синтезе дескрипторы могут играть очень важную роль, открывая существенные стороны рассматриваемого объекта, вскрывая существенную информацию о рассматриваемой системе. Эта важная информация об исследуемой или конструируемой системе может быть получена, если мы в состоянии ответить на вопросы о трех аспектах этой системы, называемых системными дескрипторами: концепт, структура и субстрат (см., напр., [7 - 9]).

Исходным дескриптором является концепт (от слова «концепция», лежащая в основе определения системы). В частном случае это может быть цель, которую реализует система. В зависимости от концепта находится то или иное строение – структура системы. И, наконец, третий дескриптор – субстрат определяет материал, из которого система «сделана».

Здесь следует подчеркнуть существенность именно данного порядка перехода от одного дескриптора к другому. Исходным дескриптором при построении, при синтезе системы, как уже отмечалось, является концепт. В зависимости от выбора концепта мы получаем разные системы. Концепт может быть задан в виде некоторого отношения, и тогда он называется реляционным, или в виде некоторого свойства, тогда имеем атрибутивный концепт.

Концепт является исходным дескриптором, поскольку остальные элементы системного рассмотрения уже предполагают некоторый концепт. Концепт же, в свою очередь, ничего не предполагает. Он задается, в основном, на основе практических потребностей. Система реализует ту или иную потребность, выраженную с помощью концепта.

Структура является следующим дескриптором системы. Обычно, если концепт имеет реляционный характер, то структура атрибутивна, представляя собой набор свойств, удовлетворяющих данному концепту. Наоборот, если концепт атрибутивен, то структура реляционна, будучи отношением, реализующим те свойства, которые выражены в концепте.

Субстрат является третьим дескриптором, поскольку он непосредственно зависит от второго дескриптора – структуры, и лишь косвенно, через структуру – от концепта. Отношение концепта к субстрату опосредовано структурой.

Это обстоятельство представляется очень важным в рамках общей теории систем и внутренней логики задач системного синтеза. Если бы не было такого опосредования, то каждый элемент субстрата обладал бы всеми признаками концепта, и в таком случае мы имели бы не систему, а понятие о концепте. И вместо субстрата системы мы получили бы объем этого понятия. Содержанием же понятия будет концепт. Таким образом, системный подход выродился бы в обыкновенную понятийную логику. Иное дело, когда субстрат непосредственно связывается со структурой. Тогда на его основе образуется целый ряд понятий, каждое из которых соответствует определенному элементу структуры. И все эти элементы связываются одним фиксированным концептом. В этом случае мы получаем не просто понятие, а систему, которая строится, конструируется «сверху» в последовательности: концепт, структура, субстрат.

## 6. О возможных разновидностях синтеза

Выше, в параграфе 3, на рис. 3 были рассмотрены определенные шаги синтеза, их некоторая последовательность, тот порядок, следуя которому целесообразно, собственно, проводить или моделировать синтез исследуемого объекта.

Теперь представляет существенный интерес: как можно было бы охарактеризовать различные виды синтеза, как провести их содержательную классификацию. Ведь в этом случае разработчики получают в свои руки определенную схему возможных решений, возможных направлений конструирования (моделирования, описания) своих объектов, получают весьма полную совокупность *возможных вариантов синтеза* рассматриваемых систем, со всем богатством сочетаний, свойств и связей. В свое время такое исследование было проделано. В работе [8] на базе языка тернарного описания (ЯТО), разработанного в рамках параметрического варианта общей теории систем для описания необходимых формализмов при работе в триаде «вещи – свойства – отношения» [9], была подробно исследована и систематизирована совокупность разновидностей синтеза (и анализа) как формального аппарата описания сложных объектов. Эта работа была проделана с использованием ряда правил, с привлечением понятий транзитивности, имплекативности, других представлений и методов алгебры логики. Здесь мы ограничимся конспективным изложением этого материала.

Используя три названных основополагающих понятия (вещи, свойства и отношения), на базе определенных логико-математических построений выводятся три *группы разновидностей синтеза*:

- a) реистический синтез, в котором ведущую роль играет первая фундаментальная компонента – вещи (res – вещь, лат.);
- b) атрибутивный синтез, в котором исходят из второй компоненты – свойство, признак, атрибут;
- c) реляционный синтез, в котором ведущую роль играют отношения (relation).

Под *реистическим синтезом* понимают объединение рассматриваемых объектов: пара сапог, микрорайон, группа студентов, макет двигателя, ведро воды, папка с файлами. Это объединение может быть реальным или мысленным; в нашем рассмотрении, где акцент делается на модельном (или проектном) представлении, имеется в виду мысленное объединение.

В языке тернарного описания рассматриваются вещи, имеющие разную степень определенности: любая (какая угодно) вещь; любая из попадающихся вещей; определенная, строго фиксированная вещь. Пример первого объекта (наиболее неопределенного): «Берешь любой кирпич...», «Задумай любое число...». Пример второго объекта: «Любой кирпич с этого склада», «Любая из отобранных боярами девушек на царские смотрины». Пример третьего, наиболее определенного объекта: «Крайняя левая опора этого моста», «Вот пуля, извлеченная из раны», «Файл “Расписание” в папке “Деканат”». Возможны и более детальные градации определенности: на смотрины бояре выбирали только боярских дочерей, и не из худородных, и не смуглых, не толстых, не худых... В этой фразе содержится также пример применения разных по смыслу ограничительных операторов «только» и «кроме».

Понятно, что объединение в систему вещей с разной степенью определенности-неопределенности приводит к разным результатам реистического синтеза. Если даже объединять в нашей модели объекты с одинаковой степенью определенности, то вдоль всей шкалы мы уже получим достаточно различающиеся результаты; тем более широкий веер результатов будет получен при синтезе объектов, обладающих разной степенью определенности. (Чрезвычайно интересно исследовать вопрос об использовании других шкал характеристики основных категорий, наряду или вместо шкалы определенности-неопределенности. Примерами таких шкал можно назвать параметр простоты-сложности, естественную и искусственную природу объектов, их гомогенный и гетерогенный характер и т. д. Разумеется, для описания градаций определенности-неопределенности в структуре языка имеется больше средств, чем, скажем, для описания естественного-искусственного, однако в рамках алгебры логики это не должно вызывать больших затруднений, а результат такого «многоаспектного» синтеза мог бы превзойти все ожидания).

Добавим, что объединение может быть не только парным, в качестве вещи может выступать множество вещей и т.д. Из сказанного видно, что типология результатов реистического синтеза весьма разнообразна.

*Атрибутивный синтез.* Эта разновидность синтеза предусматривает такое объединение двух вещей, при котором одна из них (ее называют атрибутивной компонентой) становится свойством другой (соответственно – реистическая компонента). Если мы хотим объединить такие категории, как «сотрудник Петров» (объект) и «знание компьютера» (свойство, признак сотрудника), то результатом атрибутивного синтеза будет: «Петров – продвинутый пользователь». В данном случае подсоединяемое, синтезируемое свойство (компьютерная грамотность) является по отношению к Петрову внешним, он им прежде не обладал. Но можно рассматривать синтез объекта и с таким свойством, которое было ему присуще и прежде. Мало того, атрибутивный синтез может быть представлен процессом, разворачивающимся не только во времени, но и в пространстве: американский президент с самого начала объединяет в себе пост главы государства и все атрибуты исполнительной власти; дельфины и косатки не могли бы выжить как вид, если бы не обладали очень важным свойством – коллективное спасение большого или раненого сородича, даже при прямой личной угрозе, – интересы сохранения вида важнее индивидуальных.

В некоторых работах приходится встречать выражение «понятийная структура» (например, предметной области). При ознакомлении оказывается, что на деле речь идет о «структуре понятий». Но первое является результатом атрибутивного синтеза, второе – реистического, и понятно, что эти результаты различны: под первым понимается свойство (какая структура или – какие понятия имеют место), под вторым – совокупность отношений между понятиями.

Между атрибутивным и реистическим синтезом имеется также следующее важное различие. Так как мы имеем в виду для реистического синтеза модельное (мысленное, проектное) представление, то, вообще говоря, любые объекты могут быть синтезированы реистически, поскольку их можно рассматривать совместно как некую пару объектов (близких, далеких, разноуровневых, как антиномию и т. д.), даже если не помнить первый закон Барри Коммонера: «Все связано со всем». Но для атрибутивного синтеза это не так: не *каждый* объект можно рассматривать в качестве свойства *любого* другого, если, конечно, не оперировать метафорами, используя переносный смысл: твердый характер, железная воля или патовая ситуация. В самом деле, такое свойство, как соленость, даже мысленно нельзя преставить как свойство соловьиной трели или второго члена биннома Ньютона, а твердость – как свойство телевизионной картинки. (Аристотель говорил, что бессмысленно рассуждать о слепоте или зрячести камня, поскольку этими признаками могут обладать лишь живые существа). Итак, в атрибутивном синтезе возникает необходимость запрета некоторых комбинаций. Эта проблема в ЯТО решается введением понятия «невозможная вещь» и соответствующих правил его применения. Такой прием в методологическом отношении значительно обогащает возможности атрибутивного (и реляционного) синтеза.

*Реляционный синтез.* С этим видом синтеза мы имеем дело тогда, когда одна вещь (реляционная компонента реляционного синтеза) устанавливается в качестве отношения к другой вещи (реистическая компонента реляционного синтеза). Если мы хотим объединить такие «вещи», как «группа предприятий» и «сотрудничество», то результатом такого реляционного синтеза будет: «комплементарная группа предприятий» или при более сильной связи (более выраженном отношении): «группа сотрудничающих предприятий», кластер. Очень важно отметить, что, формулируя правила реляционного синтеза, можно активно использовать основное свойство отношений, отличающее отношения от свойств: способность отношений образовывать новые объекты из тех, между которыми они устанавливаются. Примеры: супружеская пара, промышленное предприятие, рота солдат, пирамидальная сеть, рынок, электрическая розетка, миллион примеров.

В цитируемой работе ([8]) в результате рассмотрения ряда аксиом и теорем (как это было сделано и для других видов синтеза), применительно к реляционному синтезу представлен ряд выводов и ограничений. Вот несколько примеров: из данных вещей или с помощью данного отношения нельзя получить любую вещь, в определенной вещи невозможно установить любое отношение, имеет место асимметрия отношений между вещами и отношениями и некоторые другие.

В рамках *реляционного анализа* зафиксировано, например, положение о том, что субстрат определенной вещи нельзя отождествлять с самой вещью: та же самая вещь может быть реализована на разных субстратах, если сохраняется отношение, отождествляемое с самой вещью. Вообще изучение области отношений системных дескрипторов, с одной стороны, с разными конфигурациями разновидностей анализа и синтеза, с другой, – представляется чрезвычайно интересной задачей с далеко идущими следствиями и в методологическом, и непосредственно в прикладном плане.

Результаты проведенного рассмотрения иллюстрируются рядом применений к задачам науковедения, задачам анализа, синтеза и типологии знаний. Например, одно из положений исследования определяет тип знания, которое устанавливает точно фиксированные отношения в неопределенных вещах и отличные от этих отношений свойства. К такому типу знания относится, например, кибернетика, изучающая фиксированные отношения – управления и контроля в неопределенных объектах и также исследует некоторые свойства, отличные от управления и контроля, например, оптимальность. Один из вариантов реистического анализа будет соответствовать анализу субстрата, к которому относится знание. Это, в свою очередь, соответствует тому, как кибернетику делят на промышленную, техническую и т. д. Но такое членение не затрагивает того, что является существенным в этом знании. В другом варианте можно провести членение кибернетики на главы, каждая из которых охватывает отдельные части процесса управления или, в третьем варианте, разные виды управления (например, дискретное – непрерывное).

Но более интересным может быть применение атрибутивного и реляционного анализа. Их назначением будет уже не классификация, а углубление, развитие научного знания. При этом из данной структуры знания могут быть получены другие, производные от нее. Далее, с помощью реляционного синтеза может быть представлен тот тип синтеза, который реализует «пронизывание» ряда частных естественных наук более общими науками (например, методология или математика), которые отражают какую-то общую сторону (количественную, структурную, управление, ...). Такая стержневая наука представляет собой определенный тип отношений, устанавливаемых в других науках.

Заканчивая этот краткий обзор разновидностей синтеза и анализа, выскажем в качестве простого методологического примера небольшое, но существенное соображение. Даже при первом знакомстве с реляционным синтезом и реляционным анализом становится понятным достаточно спокойное отношение в серьезных публикациях последних лет к методу анализа иерархий, хотя в прежнее время он вызывал очень большой энтузиазм. Для многих разработчиков стало понятным: это действительно эффективный метод, но лишь для систем, в которых имеют место иерархические отношения. А как быть с системами, структура которых имеет гетерархический или анархический характер, или для которых «вертикальный» или «горизонтальный» характер связей вообще не важен? А как быть с системами, структура которых имеет в принципе полевой, средовой характер (или ее удобно так представить)? Ведь в настоящее время все больше систем понимаются и рассматриваются именно как полевые.

Если разработчик, занимаясь сложной системой, применяет к ней тот или иной конкретный метод, например, метод анализа иерархий, хотя бы он его и хорошо знал (или – «все так делают»), а затем удовлетворенно считает свою работу выполненной, то это порой напоминает ситуацию с поисками потерянных ключей под фонарем, потому что там светло. На деле мир отношений гораздо богаче, вряд ли он по своему разнообразию уступает миру свойств.

Из приведенного краткого обзора видно, что достаточно высокое разнообразие категорий и их типов порождает весьма большое разнообразие результатов системного синтеза, что представляется ощутимо важным для разработчиков. Эти результаты и методики подробно рассмотрены в [8]. В этой работе, как говорилось, рассмотрены также и названные выше разновидности системного анализа (реистический, атрибутивный и реляционный анализ), со своими особенностями, богатыми возможностями и результатами. Очень важно учитывать диалектические отношения между синтезом и анализом.

---

## 7. Некоторые цели и задачи системного синтеза (вместо заключения)

---

Представляет интерес провести хотя бы краткое рассмотрение функционального назначения таких важных процедур, как системное моделирование и системный синтез. Каковы их задачи, чему способствует и что продвигает системный синтез, кроме решения своих прямых задач проектирования, моделирования и построения новых сложных объектов? Отметим здесь лишь некоторые важные для нас аспекты: методологический, познавательный и некоторые другие.

Итак, применение методов и идеологии системного синтеза (и системного анализа) дает возможность обеспечить в более полном объеме и с более высоким качеством:

- объяснение (в том числе – на модели) рассматриваемых сложных процессов, явлений, ситуаций;
- формирование междисциплинарного языка высокой степени общности;
- выявление *структуры* своих знаний (т. е. лучше познать познанное – известно, что прочитать книгу – еще не значит ее познать);
- формирование своих знаний как операциональных, работоспособных, целеориентированных;
- указание возможных путей формирования новых знаний (выявление родо-видовых отношений, переход от плоских к объемным представлениям, системные модели – шаг к ликвидации разрыва между «известным» и познанным, нахождение новых имен – семиотическая линия, вакантные места в классификационных схемах,..);
- формирование (организация) контекстных отношений, формирование (организация) смыслов;
- построение моделей (больше моделей – больше точек зрения, модели как инструмент познания и деятельности, мыследеятельности);
- формирование операционального поведения;
- обучение (дать удочку, а не рыбу);
- вписывание своих знаний в общую картину мира, в широкую ноосферную парадигму;
- обучение в квадрате: «Учитель! Выучи ученика, чтобы потом было у кого учиться!» и так далее...

Некоторые вопросы синтеза, относящиеся к последним этапам приведенного перечня (в параграфе 3 настоящей работы), были рассмотрены ранее в ряде публикаций. Это касается задачи агрегирования и получения обобщенных оценок (см., напр., [4, 10, 11]), классификации и принятия решений (см., напр., [5, 11, 12]), построения Интеллектуального Агента для представления и интерпретации знаний и учета контекста [2]. Существенный материал по системному моделированию сложных объектов содержится в работах [13 – 15 и др.]. Материалы перечисленных публикаций (и ряда других) в применении к задачам системного анализа и синтеза предполагается использовать при продолжении настоящей работы.

---

## Благодарности

---

Настоящая работа была выполнена при поддержке интернационального проекта ITHEA XXI Института информационных теорий и их приложений FOI ITHEA и Ассоциации ADUIS Украина (Ассоциация разработчиков и пользователей интеллектуальных систем).

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA ([www.ithea.org](http://www.ithea.org)) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine ([www.aduis.com.ua](http://www.aduis.com.ua)).

---

**Литература**

---

1. А. Крисиллов, Е. Соловьева, А. Уемов. Краткий методологический меморандум – ч. I. Information Science & Computing. International Book Series: Knowledge – Dialogue – Solution, N. 15. – ITHEA, Sofia, 2009.
2. Крисиллов А. Системный анализ с точки зрения системологии. // «Наукові записки Міжнародного Гуманітарного Університету», № 17. – Одеса: МГУ, 2010.
3. М. Ф. Бондаренко, Е. А. Соловьева, С. И. Маторин. Основы системологии. – Харьков: ХТУРЭ, 1998.
4. Крисиллов А., Крисиллов В. Формирование целеориентированной векторной модели для построения агрегированных оценок сложных объектов. // Методы решения экологических проблем. Под ред. проф. Л. Мельника. – Сумы: ОАО «СОТ»; изд. «Козацький вал», 2005.
5. А. Д. Крисиллов. Модельное описание процессов развития: механизмы, структура, система целей, индикаторы. Decision Making and Business Intelligence Strategies and Techniques (International Book Series “Information Sciences and Computing”, N 3), Supplement to International Journal “Information Technologies and Knowledge”, v. 2/2008.
6. В. В. Налимов. На грани III тысячелетия: что осмыслили мы, приближаясь к XXI веку. – М.: «Лабиринт», 1994.
7. А. И. Уемов. Системные аспекты философского знания. – Одесса: «Негоциант», 2000.
8. А. И. Уемов. Формальные аспекты систематизации научного знания и процедур его развития. В сб. «Системный анализ и научное знание». – М.: «Наука», 1978.
9. Уемов А. И. Системный подход и общая теория систем. – М.: «Мысль», 1978.
10. А. Д. Крисиллов. Интегральная оценка социально-экономич. развития приморских областей Украины: ресурсы, ситуация, приоритеты // Національні і регіональні особливості реформування соціально-економічних відносин і регулювання екологічних процесів в Україні та Польщі.- Киев – Одесса – Варшава, 1997.
11. А. Д. Крисиллов и др. Применение квалиметрических моделей при решении социально-экономических задач. International Journal “Information Theory & Application”, vol. 10, 2006, ITHEA, Sofia, 2006.
12. А. Д. Крисиллов. Принятие решений при зависимых признаках. Proc. of International Conference “Knowledge – Dialogue – Solution” – 2001, SPb. – 2001.
13. А. И. Катренко. Системный анализ объектов та процесів комп'ютеризації. – Львів: „Новий світ – 2000”, 2003
14. Бондаренко М. Ф., Соловьева Е. А., Маторин С. И., Ельчанинов Д. Б. Системологическая технология моделирования информационных и организационных систем. – Харьков: ХНУРЭ, 2005. б) Е. А. Соловьева. «Системологическое моделирование бизнеса для разработки информационных систем». Proc. XV International Conference “Knowledge – Dialogue – Solution”, vol. # 1, June 20 – 25, 2009, Varna (Bulgaria) – ITHEA, Sofia, 2009.
15. Kr. Markov, Kr. Ivanova, I. Mitov. a) Basic Structure of the General Information Theory. IJ ITA, vol. 14, No 1. – ITHEA, Sofia, 2006. b) General Information Theory Basic Formulations). – FOI COMMERCE, Sofia, 2003.

---

**Информация об авторах**

---

**Анатолий Крисиллов** – Одесская государственная Академия холода, к. т. н., доц. кафедры информационных и коммуникационных технологий; ул. Дворянская, 1/3, Одесса - 82, 65082, Украина; тел. (0482)-632-598; моб. (38097)-291-33-24; E – m: [adkrissilov@list.ru](mailto:adkrissilov@list.ru)

**Авенур Уемов** – Одесский Государственный университет им. И. И. Мечникова, профессор кафедры философии естественных факультетов, член Академии истории и философии науки (Калифорния, США); ул. Дворянская, 2, Одесса-26, 65026, Украина; тел. (0482)-636-817.

## МОДЕЛЬ УНИВЕРСУМА

Мержвинский Анатолий

**Аннотация:** Предложена концепция построения, компоненты и структура модели универсума применительно к физической и компьютерной реализациям. Модель отображает категории и природу взаимодействий материальных и информационных объектов.

**Ключевые слова:** УНИВЕРСУМ, ОБЪЕКТ, ВЗАИМОДЕЙСТВИЯ, ОБРАЗ, КОММУНИКАНТ

**ACM Classification Keywords:** Philosophy and Methodology of Informatics

---

### Введение

Научная картина мира сложная и характерна для каждого индивидуума, так как складывается из представлений о различных предметных областях (ПО). Представление ментальных структур, изоморфных связям и отношениям между фрагментами реального мира, в машинной форме – основная задача создания баз знаний [1]. Успешное проектирование проблемно-ориентированных систем увело, на наш взгляд, исследователей от проблем создания и визуализации модели универсума (который будем рассматривать как наивысшую ступень объединения ПО). Хотя попытки "увязать картину мира во что-то целое, что подвергается конструктивному анализу и пригодное для практического использования" были сделаны ранее и выполнена декомпозиция ноосферы на «физикал», «ментал» и «культурал» [2]; распространено представление мира в виде онтологии [3] либо глобального графа, однако крупноблочной модели мира верхнего уровня, которая позволяет объединить и наглядно отобразить основные категории бытия, до сих пор не разработано. Отметим, что существующие физические теории материального мира и логические теории представления знаний, развиваясь относительно независимо, на наш взгляд, недостаточно отражают природу информационных объектов, их связь с материальными объектами. Совершенствование технологий микро- нано- оптоэлектроники и голографии позволяет по-новому подойти к визуализации данных и знаний о трехмерных, а в общем случае и  $n$ -мерных сущностях объектов с помощью трехмерных цветных отображений.

---

### Объектологический подход к представлению макромира

С точки зрения "наивного реалиста", позиции которого будем разделять в данной статье, Вселенная представляет собой совокупность объектов. Для специалиста же квантовой физики это подвижная единая ткань, которая составляется со всплесков энергии, и ни одна из частей этой "ткани" не существует независимо от целого. Однако в такой картине мира существует достаточно противоречий, которые отвлекают от созданного большинством макрообраза мира. Будем исходить из того, что при морфологическом описании материальных объектов обычно выделяют материальные, энергетические и информационные связи. Для учета взаимодействий на физическом уровне будем считать, что важнейшим свойством материальных объектов является свойство обмениваться со средой  $V$  и другими материальными объектами веществом  $G$ , излучением  $\Gamma$  разной природы, энергией  $E$  и импульсом.

**Непосредственные взаимодействия.** Современные экспериментальные данные свидетельствуют, что существует, строго говоря, только четыре качественно разных вида взаимодействий: гравитационное, электромагнитное, сильное, слабое. Эти взаимодействия считаются фундаментальными, т.е. самыми основными, исходными, первичными. Эффект взаимодействий существенно зависит от размеров объектов и расстояния между ними.

**Опосредствованные взаимодействия.** Соответственно природе и роли при действиях одних материальных объектов на другие представим этот процесс в виде схемы на рис. 1. Носителями непосредственных взаимодействий материальных объектов, как указывалось выше, являются материальные потоки и разного типа поля. Стрелки обозначают направление воздействия.

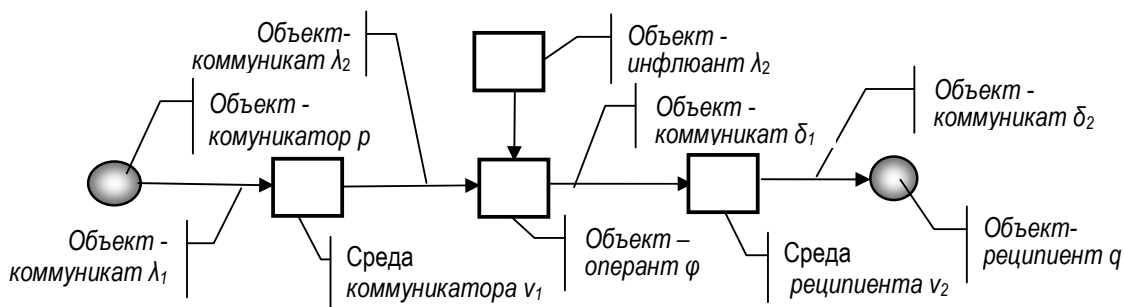


Рис.1 – Обобщенная схема опосредствованного воздействия коммуникатора на реципиента

Будем пользоваться такими определениями. **Коммуникант** - участник коммуникации, задействованный в коммуникативном акте взаимодействия между коммуникатором и реципиентом. **Коммуникатор** - участник взаимодействия, которое может порождать коммуникаты – вещественные, энергетические или полевые потоки, в общем случае потоки любой природы (материальные, денежные, информационные). **Коммуникат** - вещественный, энергетический или полевой потоки, которые способны осуществлять материальное или информационное влияние на реципиента. **Реципиенты** – характеризуются состоянием и совокупностью реакций (функцией). **Операнты** - материальные объекты, которые обеспечивают опосредствованное воздействие объектов - коммуникаторов на объекты-реципиенты; реализуются в виде машин, людей, человеко-машинных систем или систем, которые могут содержать и естественные объекты. **Объекты-инфлюанты** - управляющие, координирующие или иным образом влияющие на объекты-операнты. **Среда** - то, что окружает систему и оказывает на нее влияние.

**Классы и ипостаси материальных объектов.** Применительно к происхождению множество материальных объектов  $R$  формально опишем тройкой  $R = \langle F, W, \Omega \rangle$ , где  $F$  - класс материальных неживых естественных объектов;  $W$  - класс материальных неживых искусственных объектов;  $\Omega$  - класс материальных живых объектов.

Материальные объекты могут рассматриваться в ипостаси носителей информационных объектов (ИО). В данной работе, в зависимости от вида операции взаимодействия с другими объектами, ИО из множества  $I$  универсума  $U$  могут:

- отображать некоторую сущность (идею, идеальный объект) других объектов (прообразов)  $U$ ;
- представлять самостоятельную сущность (идею, идеальный объект), которая может непосредственно или через объекты-операнты некоторой конфигурации  $\Phi$  влиять на третьи объекты  $Q$ .

В первом случае, как отмечалось выше, это может быть "образ" объекта, а в более сложном, "модель" объекта. В случае же самостоятельной сущности это "код", "программа" или "модель".

Формально множество объектов  $I$  представим тройкой  $I = \langle X, \Theta, \Psi \rangle$ ,  $X$  - класс элементов  $x$ , в котором элементы  $x$  - образы сущностей материального мира  $R$ , полученные с помощью инструментов (фото, показание измерительных приборов и др.),  $\Theta$  - класс, в котором  $\theta$  - продукты умственной деятельности, которые зафиксированы на любом носителе,  $\Psi$  - класс, в котором  $\psi$  - ментальные объекты в живом организме.

### Акты взаимодействий, операций и процессы в ПО

**Материальные процессы** – будем рассматривать как процессы изменения структуры и состояния объекта, обмена материального объекта  $g$  с внешней средой коммуникатами (вещественными потоками  $g$ , полевыми  $\gamma$ , энергией  $e$  или импульсом). Энергетические и вещественные взаимодействия объектов будем рассматривать как симметричные, т.е. сколько один отдал столько же другой или другие получили. Движение объектов в среде  $V$  и времени  $T$  могут отображаться в акте  $o$  элементарного действия коммуникатора  $p$  на реципиента  $q$ , например, в виде переходных процессов (линейных или нелинейных



траекторий изменения характеристик, в которых может учитываться действие обратной связи между реципиентом и коммуникатором) или операций  $O$ , в которых учитывается опосредствованные взаимодействия между реципиентом и коммуникатором, а также некоторого процесса  $\pi$ , как последовательности операций.

Таким образом, взаимодействия объектов - это процессы влияния коммуникаторов  $P$  на реципиентов  $Q$  с помощью коммуникатов  $\Lambda$  и наоборот, с учетом их поведенческих свойств, особенностей и взаимной обусловленности распространения в среде  $V$  конкретных коммуникатов  $\lambda_P$  и  $\lambda_Q$ , где  $\lambda_P$  - коммуникаты объектов  $P$ , и  $\lambda_Q$  - коммуникаты объектов  $Q$ .

*Информационные процессы* будем рассматривать как процессы формирования образов сущностей материальных или информационных объектов универсума  $U$  в некотором материальном носителе, их обработки, распознавания, идентификации, накопления в информационной среде, процессы опосредствованного действия информационных объектов из множества  $P_i$  на материальные объекты-реципиенты из множества  $Q$  с помощью соответствующих объектов-оперантов из множества  $\Phi$ , а также поиска, распространения, формирования и формализации знаний человеком.

В зависимости от рассматриваемой ипостаси коммуникаторов и реципиентов - как материальных  $R$  или как информационных  $I$  (носителей или идеальных) - в акте взаимодействия могут принимать участие такие комбинации коммуникантов:  $R-R$ ,  $R-I$ ,  $I-I$ ,  $I-R$ . Эти комбинации определяют категории *актов* взаимодействий коммуникантов (для выполнения требуется время) или *операций* (последовательности актов опосредствованных взаимодействий с обратной связью) или *процессов* (последовательности операций взаимодействий). В случае опосредствованных актов взаимодействий результат может зависеть от влияния на оперант  $\varphi$  управляющих, координирующих или иным образом влияющих объектов (объектов-инфлюантов). Объекты-инфлюанты могут включать среды, временные синхронизаторы, программы и модели желательных и реальных процессов. Таким образом, процесс может рассматриваться как объект, который существует и в пространстве и во времени.

$R-I$ -операции выполняются с помощью сенсоров и могут быть определены как "сенсорные" операции, а  $I-R$ - операции выполняются с помощью эффекторов (исполняющих устройств) и могут быть определены как "эффекторные" операции. Таким образом, в природе с помощью коммуникатов и объектов-оперантов могут иметь место процессы отображения некоторой сущности материальных объектов в информационные и, наоборот, информационных объектов в материальные.

**Особенности  $R-I$ -преобразований.**  $R-I$ -операции более широкое понятие, чем общепринятые «операции ввода – вывода». В информатике, *ввод/вывод* трактуется как взаимодействие между обработчиком информации (например, компьютер) и внешним миром, который может представлять как человек, так и любая другая система обработки информации. В нашем случае  $R-I$ -операции включают также взаимодействие коммуниката и коммуниканта. Технически технология преобразования материального объекта в его образ может быть реализована двумя способами:

- *Анализ с помощью технического средства* (соответствующего сенсора) *первичного коммуниката*  $\delta$  (например, электромагнитного излучения или вещественного потока), которые создаются и эммитируются коммуникатором (рассматриваемого как активный объект);
- *Зондирование* коммуникатора некоторым влияющим коммуникатом  $\Lambda$  (который бомбардирует пассивный объект - коммуникатор) и анализ с помощью соответствующего сенсора вторичного коммуниката  $\delta$ , (который эммитируется коммуникатором).

**$I-I$ - преобразование** – известная основная операция теории связи; например, передача изображения из одного компьютера на другой.

---

### Отображение множества материальных объектов ПО

---

Концепция отображения объектов и взаимодействий в предлагаемой модели исходит из возможности отображения их сущности в определенных материальных средах в виде композиции отображающих

каморок (объектов-носителей), которые могут содержать поверхностные (двумерные пиксели), объемные (воксели) и объемно-пространственные ячейки (имеющие внутреннюю структуру) и допускающие запись в них и считывание ссылок к внешней памяти. На фиг. 2 приведена схема расположения каморок, позволяющая отобразить объекты на рис. 1 (опосредствованное влияние  $i$ -го коммуникатора множества  $P$  на  $j$ -й реципиент множества  $Q$ ).

**Формализация отображения ПО.** Структурные элементы схемы на рис. 2 могут быть положены в основу онтологического описания взаимодействующих на физическом уровне объектов ПО. При таком подходе множество материальных объектов  $R$  ПО формально может быть представлено пятеркой  $R = \langle P, Q, \Lambda, \Delta, \Phi \rangle$ , где  $P$  - множество объектов-коммуникаторов, которые являются источниками формирования множества объектов-коммуникатов  $\Lambda$ ;  $Q$  - множество объектов-реципиентов, на которые воздействуют объекты-коммуникаты  $\Delta$ ;  $\Phi$  - множество объектов-оперантов, которые принимают участие в формировании вещественных или полевых действий на объекты-реципиенты из множества  $Q$ . Объекты-коммуникаты  $\Lambda, \Delta$  - есть носители ввоздействий: это вещественные потоки  $G$  (характеризуются массой покоя) и физические поля  $\Gamma$ , которые в общем случае могут характеризоваться потенциальной и кинетической энергией  $E$ .

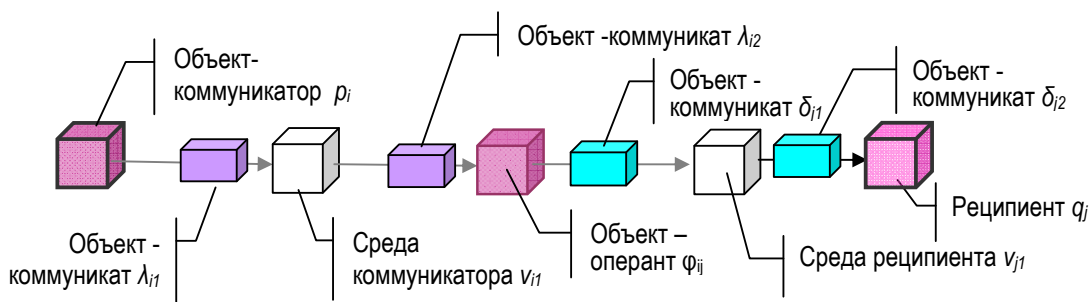


Рис.2. Схема соединений каморок для отображения действия коммуникатора на реципиента

### Концепция отображения взаимодействующих объектов универсума

При наличии  $n$  классов объектов  $R$  количество вариантов связей классов возрастает пропорционально  $n^2$ , поэтому модель на рис. 2 становится не наглядной. Кроме того, иногда важным есть взаимодействие с объектами окружения ПО - миром внешней действительности. Для отображения универсума предлагается ячеистая (каморковая) модель которая может быть реализована в материальных или виртуальных средах на основе каморок отображения.

**Принцип построения ячеистой модели универсума** состоит в выделении для каждого объекта  $u_i$  универсума  $U$  каморки объекта  $y_i$  из множества  $Y$ , формировании в каморке объекта  $y_i$  отображения сущности объекта или ссылки на внешнюю память, объединению каморок упорядоченных по категориям и ролевым признакам отображаемых объектов в структурные элементы - модули, например, в форме параллелепипедов, и формировании отображения универсума объектов в виде композиции модулей  $Y_\alpha$ , где  $\alpha$  - множество индексов модулей [4].

Модель может быть визуализирована в виде композиции каморок (рис. 3), содержащих отображающие объекты - носители - поверхностные, объемные и объемно-пространственные ячейки. Расположение и доступ к каморке определяется координатами  $x, y$  и  $z$  в Декартовой системе координат и не зависит от того, где физически находится ее содержание - в теле блока отображения, или на внешнем носителе.

### Визуализация модели отображения в форме куба

**Расположение модулей.** Ячеистая модель отображения универсума выполнена согласно рис. 3 в форме куба в виде композиции восьми модулей. Расположение модулей привязано к правой системе Декартовых координат. Измерения объектов визуализируются в направлениях осей  $X, Y$  и  $Z$ , а именно:  $Y$  - ось

измерений категорий коммуникаторов;  $X$  - ось измерений категорий реципиентов;  $Z$  - ось измерений классов коммуникантов.  $m$ ,  $n$  - соответственно объекты или классы взаимодействующих коммуникаторов и реципиентов. Члены измерений визуализируются как точки или звенья, которые откладываются на осях модуля. Состав и функциональное назначение модулей следующие:

Модуль 1 (III октанта) содержит камеры объектов-коммуникаторов множества ( $Y_P$ ). Графические элементы модуля 1 на грани 1 (рис. 3, фото 1) расположены, таким образом, что последовательность категорий объектов-коммуникаторов направлена по оси  $-Y$ , а в направлении оси  $Z$  отображаются классы, подклассы, абстрактные и/или конкретные объекты-коммуникаторы.

Модуль 2 (III октанта) содержит камеры исходных  $Y_L$  коммуникаторов  $P$ . Примеры коммуникаторов: фотонный поток, электронный поток, радиоволна (в общем случае электромагнитная волна), лист бумаги, конверт, посылка почтовая или контейнер транспортных перевозок.

Модуль 3 (III октанта) содержит камеры среды  $Y_{VL}$  для отображения пространства объектов среды, в котором распространяются исходные коммуникаты  $L$  между коммуникаторами  $P$  и входными элементами объектов-оперантов  $\Phi$ .

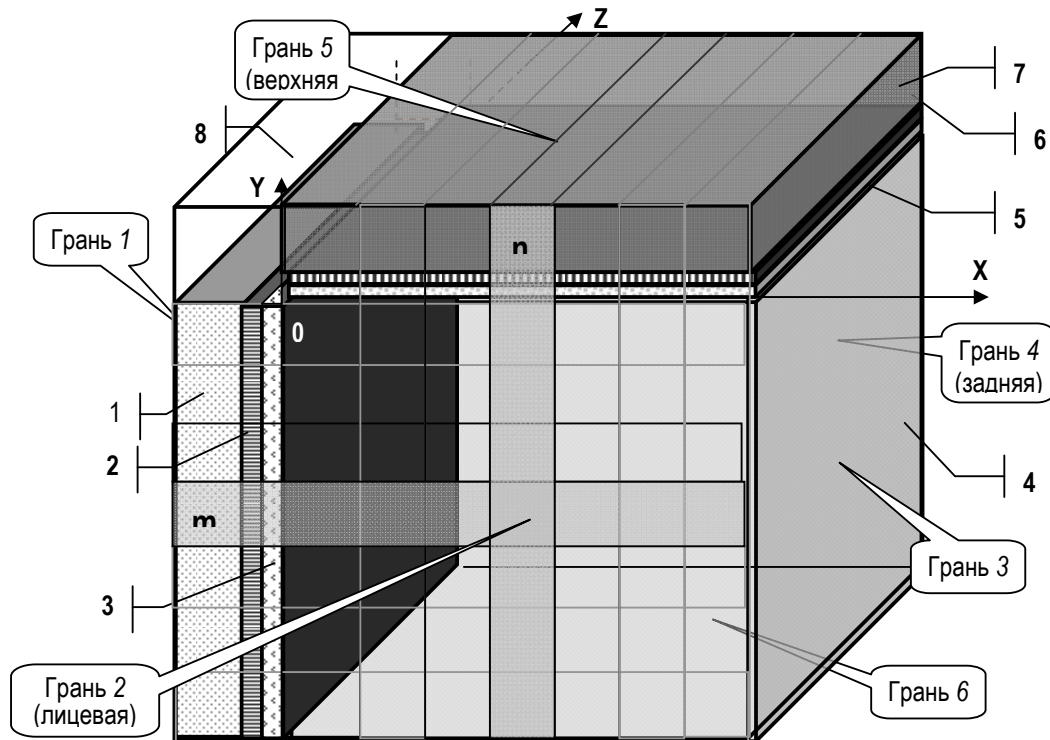


Рис. 3. Структура модели в форме куба.

Модуль 4 (IV октанта) содержит камеры объектов-оперантов  $Y_\Phi$  для отображения объектов-оперантов  $U_\Phi$ , с помощью которых осуществляется взаимодействие объектов-коммуникаторов и объектов-реципиентов. Сущность объектов-оперантов заключается в превращении материальной ( $g_L, \gamma_L$ ) и энергетической  $e_L$  компонент коммуниката  $L$  в материальные ( $g_\delta, \gamma_\delta$ ) и/или энергетические ( $e_\delta$ ) компоненты коммуниката  $\Delta$ . Учет возможности на входах и выходах разных комбинаций  $g, \gamma$ - и  $e$ -компонент коммуникаторов в общем случае создает достаточно много вариантов (64) превращений коммуникаторов. В практике конкретные вещественные  $G$ , полевые  $\Gamma$  или энергетические  $E$  составляющие или их комбинации есть доминирующие, а другие могут быть неактуальными. С учетом этого выделим такие категории операций объектов-оперантов:

- $r-r$  - операции, сущность которых определяется превращением материальной ( $g_L, \gamma_L$ ) и/или энергетической  $e_L$  составных коммуниката  $L$  в материальные ( $g_\delta, \gamma_\delta$ ) и/или энергетические ( $e_\delta$ )

компоненты коммуниката  $\Delta$ . Операции синтеза и декомпозиции сложных операций  $r$ - $r$  категории могут быть описаны методами формальных технологий [5].

- $i$ - $i$  - операции, в которых входной информационный объект множества  $I$  (или поток объектов), преобразовывается в выходной также информационный объект (или поток информационных объектов). Операции выполняются материальными объектами-оперантами (аппаратными, программно-аппаратными, программными, а также биообъектами, которые могут создавать информационные среды); на верхнем уровне обычно описываются логико-математическими моделями, которые не включают физические преобразования материальных носителей информационных объектов.
- $i$ - $r$  операции, в которых исходный информационный объект множества  $I$  (или поток объектов), независимо от носителя, в конечном счете превращается в выходной материальный и/или энергетический объект (или поток объектов). Операции выполняются исполнительными устройствами (эффекторами, которые с помощью драйвера исходного устройства реализуют действие информационного объекта на материальный объект или систему) и агрегатами в более сложных случаях.

Трактовка  $r$ - $i$  операций - очевидна. В простейшем случае выполняются сенсорами (формирование образов сущностей материальных объектов).

Информационные среды, в которых выполнение  $i$ - $i$ ,  $r$ - $i$  и  $i$ - $r$  операций реализуется в виде искусственных твердых конструкций, в практике называются «аппаратная часть». Для общности, например, со случаем использования программно-аппаратных средств или радиоволн в ионосфере,  $i$ - $i$ ,  $r$ - $i$  и  $i$ - $r$  преобразователи, при рассмотрении их в роли объектов-коммуникантов, в данной работе имеют обобщенное название  $r$ -информ-объекты или  $ri$ -объекты.

Модуль 5 (I октант) содержит камеры среды  $Y_{VD}$  для отображения пространства объектов среды между исходными элементами объектов-оперантов  $\Phi$  и реципиентов  $Q$ , в котором распространяются действующие коммуникаты  $\Delta$ , (в простейшем случае содержит камеры пространственных интервалов).

Модуль 6 (I октант) содержит камеры действующих коммуникатов  $Y_{\Delta}$ . В этих камерах отображаются объекты-коммуникаты  $\Delta$ , которые воздействуют на объекты-реципиенты.

Модуль 7 (I октант) содержит камеры объектов-реципиентов  $Y_Q$ . Модуль 7 расположен ортогонально модулю 1 и графические элементы грани 5 расположены так, что подмодули камерок последовательности категорий объектов-реципиентов направлены по оси  $X$ , а классы, подклассы и конкретные объекты-реципиенты могут отображаться в камерах в направлении оси  $Z$ .

Среда, в котором происходят протяженные процессы взаимодействия объектов, движение объектов в пространстве  $V$  и времени  $T$ , переходные характеристики процессов изменения состояния объектов, их метрика отображаются в камерах  $Y_{STP}$  модуля пространственно-временных интервалов и процессов  $\delta$ .

---

### Описание графики граней физической реализации модели

---

На фото 1 приведен пример физической реализации модели универсума в виде куба. Графические элементы поверхностных ячеек на гранях модели определяется проекцией на грани соответствующих подмножеств камерок. Эти проекции приведены на развертке граней 1,2,5 (рис. 4) и могут отображать иерархическую структуру и взаимосвязи объектов универсума  $U$ .

Множество объектов  $R$  наряду с подмножествами неживых объектов  $F$ , живых -  $\Omega$  и искусственных объектов  $W$  также включают подмножество систем  $S$ , которые могут представлять определенную совокупность конфигураций элементов разных категорий. Живые объекты  $\Omega$  - могут рассматриваться на разных уровнях: молекулярном, субклеточном, клеточном, органно-системном, организменном и других. Поэтому этот термин достаточно широкий и не однозначный. Объем понятия включает подмножество множества биологических объектов  $B$ . Биологические объекты  $B$  включают категорию «человек»  $H$ .

Искусственные объекты  $W$  представлены категорией не информатизированных объектов  $D$  (d-объектов) и категорией  $K$  ( $r$ -информ-объектов), которые могут оперировать также с информационными объектами категории  $I$ .

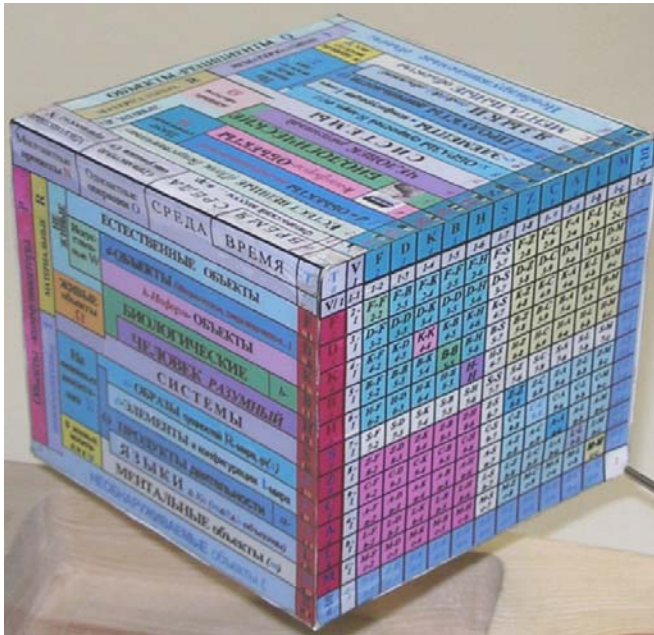


Фото 1 Общий вид модели в форме куба

В множестве информационных объектов  $I$  выделено подмножество ментальных объектов  $\Psi$  и подмножества информационных объектов на неживых носителях  $\Sigma$ .  $\Psi$  - подмножество ментальных объектов, порожденных в индивидууме или коллективе, живых организмах в результате познавательной деятельности (образы внешнего мира, идеальные объекты (эмпирические и теоретические (конструкты, абстрактные объекты)), сигналы, конфигурации, мысли, базовые когнитивные операции, ...). Подмножество  $\Sigma$  (информационные объекты на неживых носителях) включает подмножество  $X$  - образы сущностей  $R$ -мира и подмножество  $A$  - продукты деятельности информационного мира. Категория объектов  $X$  включает множество  $X_1$  - информационных объектов, которые отображают некоторые сущности

$R$ -мира -  $x'$ -образы, сформированные материальными средствами без участия оператора, и множество  $X_2$  -  $x''$ -образы, представляющие  $x'$ -образы, однако скорректированы действующим индивидуумом, например, фотопортрет, отретушированный художником соответственно его мировосприятию ( $y = \varphi(x)$ ) или формализованную модель некоторой сущности.

В категорию  $\Theta$  - продукты деятельности информационного мира объединены любые медиа-продукты умственной деятельности и зафиксированные на носителях (языка, тексты, компьютерные программы, рисунки, мелодии, фильмы, устное творчество, ...), которые доступны пользователям. Множество *Продукты деятельности*  $I$ -мира включает  $s$ -элементы и собранные из них генетические конструкции  $I$ -мира, которые могут использоваться для построения более сложных конфигураций. Элементы категории  $C$  ( $s$ -элементы) представляют первичные математические понятия (точка, линия, поверхность, объемная фигура,  $n$ -мерная фигура, число, множество, математические структуры) или отображение простейших физических сущностей материального мира (в виде точечных, линейных, объемных элементов для макромира и пространственных структур микромира - атомов, частиц и т.д.), которые моделируют свойства вещественных объектов и физических полей.

В множестве информационных объектов  $I$  выделена специфическая категория отображения объектов  $U$ : язык  $L$ . Кроме ментальной составляющей  $m$ , составляющей на неживых носителях (алфавит, ...)  $\theta$  включает также коммуникационную составляющую  $l$  в виде образцов звуковых колебаний в некоторой материальной среде и их отображении на носителе, например магнитной пленке.

Категория объектов, которые не обнаруживаются инструментальными средствами, обозначены как  $\Xi$  (ангелы, Бог, русалки, домовый, НЛО).

На грани 2 (рис. 4) нанесена категориальная сетка *камарок объектов-оперантов* в форме квадратной матрицы, поля камарок исходных и действующих коммуникатов (узкие горизонтальная и вертикальная полосы), мнемонические обозначения категорий *объектов-коммуникантов*. Мнемоника обозначения

классов объектов-операндов образованная путем соединения  $i$ -го мнемонического обозначения объекта-коммуникатора с  $j$ -м обозначением объекта-реципиента (рис. 4).

Грани 3,4 и 6 могут использоваться для иллюстраций характерных образов категорий объектов.

Отметим, что для упрощения возможности взаимодействия оператора с компьютером, оборудованным двумерным монитором, трехмерная модель отображения универсума может быть визуализирована в виде двумерной (рис. 4), полученной как развертка граней 1,2 и 5 куба.

Рис 4 Развертка граней 1,2 и 5 модели в форме куба

Многоактные процессы П		Одноактные операции О	СРЕДА	Время	ОБЪЕКТЫ-РЕЦИПИЕНТЫ Q														
					МАТЕРИАЛЬНИ R					НЕМАТЕРИАЛЬНИ I									
ЧАС	СРЕДОВИЩЕ (Фізичний вакуум... та ін.)																		
	0	1	2	3	4	5	6	7	8	9	а	б	в	г	д				
Т	V	F	D	K	B	H	S	X	C	A	L	M	Є	2					
V	V/I	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-а	1-б	1-в	1-г	1-д				
Объекты коммуникаторы Р	Материальные R	Естественные объекты													F				
		Не живые	d-объекты													D			
		Искусственные W	r-информ-объекты													K			
	Живые объекты Ω	Биологические													B				
		Человек разумный													H				
Информационные I	На неживых носителях Σ	Системы													S				
		Образы сущностей R-мира													X				
		с-элементы и констр I-мира													C				
	В живых носителях Ψ	Продукты деятельности													A				
		Языки (т- & θ- & λ- объекты)													L				
Ментальные объекты													M						
Необнаруживаемые объекты													Є						

### Активизация модулей модели универсума во времени

Возможные изменения структуры и динамики ПО при процессах, которые содержат более, чем один акт взаимодействий, могут быть отображены в виде диаграммы на рис. 5, представляющей развертку во времени процесса активизации и функционирования объектов-коммуникантов, отображаемых в камерках  $K_1-K_n$ , и объектов-операндов, отображаемых в камерках  $O_1-O_n$ . На основе развертки могут строиться каузально-эмпирические модели, модели взаимодействий верхнего уровня для уже конкретных процессов (концептуальные, графы-модели, логические), определяться такие понятия, как терминальные объекты, такты, агрегации измерений. Таким образом, на основе развертки воздействий коммуникантов могут быть сделаны эмпирические обобщения для конкретной ПО и продемонстрирован переход от физических знаний к логическим.

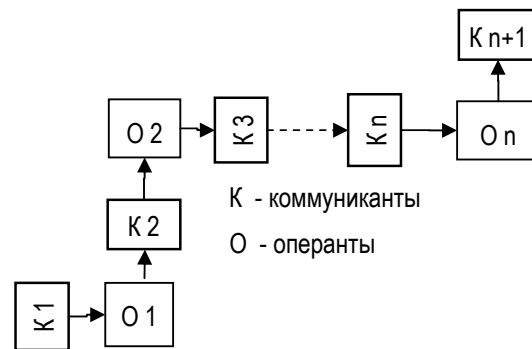


Рис. 5 Развертка воздействий коммуникантов

## Выводы

На основе важнейших поведенческих свойств материальных объектов - способности обмениваться с внешней средой веществом  $G$ , излучением, энергией  $E$  и импульсом – введены понятия «коммуникат», «оперант», «инфлюант», которые позволили визуализировать взаимосвязи объектов универсума на физическом уровне. Разработанная ячеистая модель отображает категории и природу материальных и информационных взаимодействий объектов. Модель может быть выполнена в виде трехмерной физической модели, позволяющей хранить выделенные человечеством наиболее обобщенные знания о мире (как наглядное пособие для обучения принципам классификации, устройство ввода, сувенир), и в виде программной модели компьютера для хранения более глубоких знаний. Возможное применение компьютерной модели - в составе агрегата «обобщенная компьютерная модель универсума - конкретная модель ПО» для хранения, идентификации и классификации объектов в информационных системах, работающих со знаниями.

## Литература

1. Кургаев А.Ф. Проблемная ориентация архитектуры компьютерных систем.- Киев: Сталь, 2008.- 540 с.
2. Лозовский В.С. К десигнативной теории имен. <http://vloz.mylivepage.com/wiki/155/83>
3. Палагин О.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: "УНІВЕРСУМ-Вінниця", 2005.- 680 с.
4. Патент Промышленный образец № 19543 від 12.10.2009 «КОМПЛЕКТ ПРИСТРОЇВ ДЛЯ ВІДОБРАЖЕННЯ УНІВЕРСУМА». Автор: Мержвинський Анатолій Олександрович.  
<http://base.ukrpatent.org/searchBul/search.php?action=viewdetails&IdClaim=21733&chapter=variants&dbname=pp>
5. Крылов С.М. Формальная технология в философии, технике, биоэволюции и социологии. - Самара: СамГТУ, 1997.-180 с.

## Authors' Information



**Мержвинский Анатолий Александрович** – Институт кибернетики им. В.М.Глушкова НАНУ, 03680 МСП Київ-187, пр-т Академіка Глушкова,40, Украина; старший научный сотрудник отдела 220 ИК НАНУ

e-mail: merjv@mail.ru

Major Fields of Scientific Research: физико-технологические проблемы кибернетики, микро-оптоэлектроника, биосенсорика

## СИТУАЦИОННЫЙ ЦЕНТР КАК ИНСТРУМЕНТ ДЛЯ ПРОВЕДЕНИЯ ФОРСАЙТНЫХ ИССЛЕДОВАНИЙ

**Виталий Вишневецкий, Сергей Симонов**

**Аннотация:** Рассмотрена необходимость прогнозирования научно-технического и социального развития стран с помощью форсайтных исследований. Предложена концепция использования Ситуационных центров как средства автоматизации форсайтных исследований, а также организации работы малых экспертных групп. Для этого предлагается создать веб-ориентированное решение для автоматизации различных этапов метода Дельфи - формирование группы экспертов, создание анкет, проведение опроса и генерации результатов опроса.

**Ключевые слова:** ситуационный центр, экспертные методы, форсайтные исследования, метод Дельфи.

**Ключевые слова по классификации ACM:** H.4 INFORMATION SYSTEMS APPLICATIONS.

---

### Введение

Современный мир все больше стремится к многополюсности и ни одна из развитых стран уже не может вести полномасштабные исследования по всем существующим научным направлениям. В таких условиях очевидным становится определение приоритетов научно-технического и инновационного развития с целью сохранения лидирующих позиций в исторически сложившихся наиболее развитых областях для каждой конкретной страны, а также наиболее перспективных направлений для получения дополнительных конкурентных преимуществ в будущем. Но необходимость прогнозирования изменений касается не только научно-технической, но также социальной и общественно-политической сфер. Страны, претендующие на лидерство в современном мире, уже не могут позволить себе жить так, как жилось раньше. Поэтому они активно работают над созданием своего социального будущего, включают в этот процесс общество и соответствующим образом пытаются трансформировать устаревшие институты. Каждая из этих стран реализует свои собственные программы по прогнозированию как научно-технического, так и общественно-политического развития.

Появившись более полувека тому назад Форсайт (от английского Foresight – «предвидение» или «взгляд в будущее») стал на сегодня наиболее признанным и эффективным инструментом определения стратегических приоритетов. Есть множество примеров очень успешных на наш взгляд форсайтных исследований.

Так, в Японии начиная с 1971 года результаты Форсайта определяют прогноз развития науки и технологий в стране на ближайшие 30 лет. При этом результаты «корректируются» каждые 5 лет. В настоящее время идет подготовка уже девятого прогноза.

Другой яркий пример – Великобритания, в которой методы Форсайта начали активно развиваться в начале 1990-х. Программа Форсайта в этой стране прошла уже два цикла в 1994 – 1999 гг. и 1999 – 2002 гг., а в настоящее время начат третий. Первый цикл был направлен на оценку перспектив сферы науки и технологий. Второй цикл ориентирован на решение ключевых проблем общества: стареющее население, профилактика преступности и т.д. Третий цикл затрагивает небольшое число специфических проблем, таких как киберпреступность и инфекционные болезни.

В Германии результатом форсайтных исследований в тесном сотрудничестве с японскими коллегами (Дельфи-93 и Дельфи-98) стала своя уникальная программа FUTUR (другое название – «Немецкий исследовательский диалог»), позволяющая формировать стратегическое видение для Министерства образования и науки на следующие 20 лет.

Помимо упомянутых стран форсайтные исследования проводят сейчас большинство стран-членов Европейского союза, США, Россия и многие другие страны, правительства которых осознали, что без



систематического исследования перспектив невозможно построить стратегический план необходимых изменений для развития инновационной конкурентоспособной экономики.

### Что такое Форсайт?

«Форсайт — это систематические попытки оценить долгосрочные перспективы науки, технологий, экономики и общества, чтобы определить стратегические направления исследований и новые технологии, способные принести наибольшие социально-экономические блага», - такое определение дал Бен Мартин (SPRU, University of Sussex). Мы попытаемся привести развернутое определение и рассмотрим методики проведения таких исследований.

Как мы уже упоминали Форсайт это инструмент формирования стратегических приоритетов. То есть при его помощи определяются возможные варианты будущего. Основой для оценки этих вариантов являются экспертные методы. Очевидно, что для выработки стратегических приоритетов необходима мобилизация большого количества участников. Для этого и предназначены форсайтные исследования, позволяющие при помощи привлечения «коллективного разума» достичь качественно новых результатов.

Набор методов, применяемых в том или ином форсайтном проекте, выбирается с учетом множества факторов: временных и ресурсных ограничений, наличия достаточного количества высококвалифицированных экспертов, доступа к источникам информации и т.д. Современная методология форсайт-исследований вобрала в себя десятки традиционных и достаточно новых экспертных методов. Определение нужной комбинации методов не имеет однозначного решения, но все их можно представить в виде так называемого «ромба Форсайта» (Рис.1), в вершинах которого находятся четыре наиболее важных фактора, определяющих успешность любого экспертного исследования: креативность, экспертиза, взаимодействие и доказательность [Соколов, 2007].



Рис.1 «Ромб Форсайта»

В зависимости от того, какие факторы имеют большее значение для конкретного исследования, подбирается комбинация методов расположенных ближе к соответствующим вершинам. Обычно в каждом из форсайт-проектов применяется комбинация различных методов, в числе которых экспертные панели, метод Дельфи, SWOT-анализ, мозговой штурм, построение сценариев, технологические дорожные карты, деревья релевантности, анализ взаимного влияния и т.д. Как мы уже упоминали, чтобы учесть все возможные варианты и получить полную картину привлекается, как правило, значительное число экспертов. Так, в японских долгосрочных прогнозах научно-технологического развития, проводимых каждые пять лет, участвует более 2-х тысяч экспертов, которые представляют все важнейшие направления развития науки, технологий и техники, а в последнем корейском проекте участвовали более 10 тысяч экспертов!

---

### Метод Дельфи

---

Из представленных методологий самыми популярными являются критические технологии (используется в США, Франции, России) и метод Дельфи (Япония, Германия, первый Форсайт Великобритании). Поскольку наиболее популярной методологией сейчас является метод Дельфи, и, к тому же, именно он используется для форсайтных исследований в нашей стране [Малицкий, 2004], на нем мы остановимся более подробно.

Метод Дельфи или дельфийский метод, заимствовал своё имя от Дельфийского Оракула. Методология была разработана еще в 1950-1960 годы корпорацией RAND для анализа планов атомной войны США [Helmer, 1966]. Это многоэтапный метод, предусматривающий первоначальное изолированное вынесение экспертами своих суждений и дальнейшую многократную их корректировку на базе ознакомления каждого эксперта с суждениями других экспертов до тех пор, пока величина разброса оценок не будет находиться в рамках заранее устанавливаемого желаемого интервала варьирования оценок.

Метод Дельфи относится к классу количественных методов групповых экспертных оценок. Опрос экспертов проводится в 3-4 тура, состоящих из серии анкет, вопросы конкретизируются от тура к туру. Для реализации этого метода необходимо создание группы экспертов верхнего уровня или, так называемой, малой группы экспертов, которая после каждого тура производит статистическую обработку полученной информации и на её основании «конкретизирует» анкеты для следующего тура. То есть, прежде всего, малая группа определяет область предпочтительных количественных значений объектов. После такой проверки проводится очередной тур.

Процедуру экспертного опроса по методу Дельфи можно разделить на несколько этапов.

1. Формирование малой группы экспертов

Задача этой группы заключается в организации процедуры экспертного опроса.

2. Формирование большой группы экспертов

Также ее называют группой нижнего уровня. В соответствии с методом Дельфи группа экспертов должна включать не менее 10-15 специалистов для каждого направления исследования. Компетентность экспертов определяется путем анкетирования, анализом индекса цитирования (количества ссылок на работы данного специалиста), использованием листов самооценки. Одним из наиболее распространенных способов создания группы нижнего уровня является метод снежного кома. Детально вопрос создания пула экспертов рассмотрен в работе [Малышев, 2009].

3. Составление анкет

Как известно, все виды вопросов можно типизировать. В методе Дельфи вопросы первых туров анкетирования в основном открытого или полукрытого типа (то есть предполагают возможность свободных ответов) ближе к завершению исследования вопросы все более конкретизируются (закрытый тип), то есть формулировки вопросов становятся четкими, однозначно трактуемыми и предполагают однозначные ответы. Таким образом, выполняется условие конвергентности результатов опросов.

4. Проведение экспертизы

Группа экспертов верхнего уровня проводит статистическую обработку полученной от всех экспертов информации. Для этого рассчитывается среднее значение исследуемого параметра, средневзвешенное значение исследуемого параметра, определяется медиана как средний член общего ряда чисел, полученных от экспертов и область доверительности. На основании этой информации делаются выводы по усредненным результатам и составляются анкеты следующего тура.

Эксперты должны обязательно познакомиться с результатами и выводами предыдущего тура, после чего проводится следующий. По результатам представленных расчетов эксперты нижнего уровня могут видеть, как соотносится их мнение с мнением всей группы экспертов. Они могут изменить свои мнения или оставить прежними, но в этом случае выдвинуть контраргументы в свою пользу. При этом строго соблюдается принцип анонимности. Таким образом проводится 2-3 тура. В итоге получаем довольно точную групповую оценку.

Шаги проведения опроса повторяются до тех пор, пока ответы не начнут сходиться к каким то определенным значениям с установленным ранее разбросом, то есть фактически должно появляться совпадение мнений.

#### 5. Подведение итогов опроса

На основании полученных результатов эксперты верхнего уровня строят итоговый прогноз исследования.

---

### **Применение СЦ для форсайтных исследований**

---

Как мы уже упоминали, метод Дельфи один из наиболее формализованных среди методик, применяемых для форсайтных исследований, а это значит, что он легче других должен поддаваться автоматизации. Набором информационных технологий для поддержки автоматизации разных этапов Форсайта обладают, на наш взгляд, такие комплексные решения как Ситуационные центры (СЦ) [Морозов, 2003]. Кроме того, именно СЦ известны как наиболее приспособленный инструмент выработки стратегических решений [Морозов, 2003], что и является целью большинства форсайтных опросов.

В данном контексте под СЦ мы будем понимать совокупность программно-технических средств, научно-математических методов и инженерных решений для автоматизации процессов отображения, моделирования, анализа ситуаций и управления организационными системами.

В частности из описанных выше этапов Дельфи-опроса могут быть автоматизированы процесс подбора группы экспертов нижнего уровня [Малышев, 2009], процесс анкетирования и коллективного редактирования анкет между турами опросов группой экспертов верхнего уровня, а также генерация результатов опросов всех туров исследования, включая окончательные результаты исследования.

Наше видение возможного места СЦ в форсайтных исследованиях по прогнозированию научно-технического и инновационного развития Украины схематически можно представить в виде Рис.2. В качестве примера мы взяли метод Дельфи с тремя циклами опросов в комбинации с разделением экспертов на две группы – верхнего и нижнего уровней. Именно такая методика использовалась в нашей стране для Форсайта 2004-2006 годов [4]. Как видно из Рис.2 центральную роль для автоматизации таких исследований играет сервер СЦ, обеспечивающий как процедуру web-анкетирования экспертов нижнего уровня, так и работу малой группы в пределах СЦ.

Можно выделить ряд преимуществ в использовании СЦ по сравнению с другими системами для поддержки принятия решений (СППР):

- СЦ направлены на увеличение объема информации по рассматриваемой проблеме;
- СЦ позволяют использовать такие неформализуемые знания как интуиция и опыт;
- СЦ позволяют получить конкретную информацию, недостающую на данный момент;
- СЦ порождают альтернативные варианты решений, которые можно сравнить;
- СЦ объединяют усилия целого коллектива, создавая соответствующий синергетический эффект.

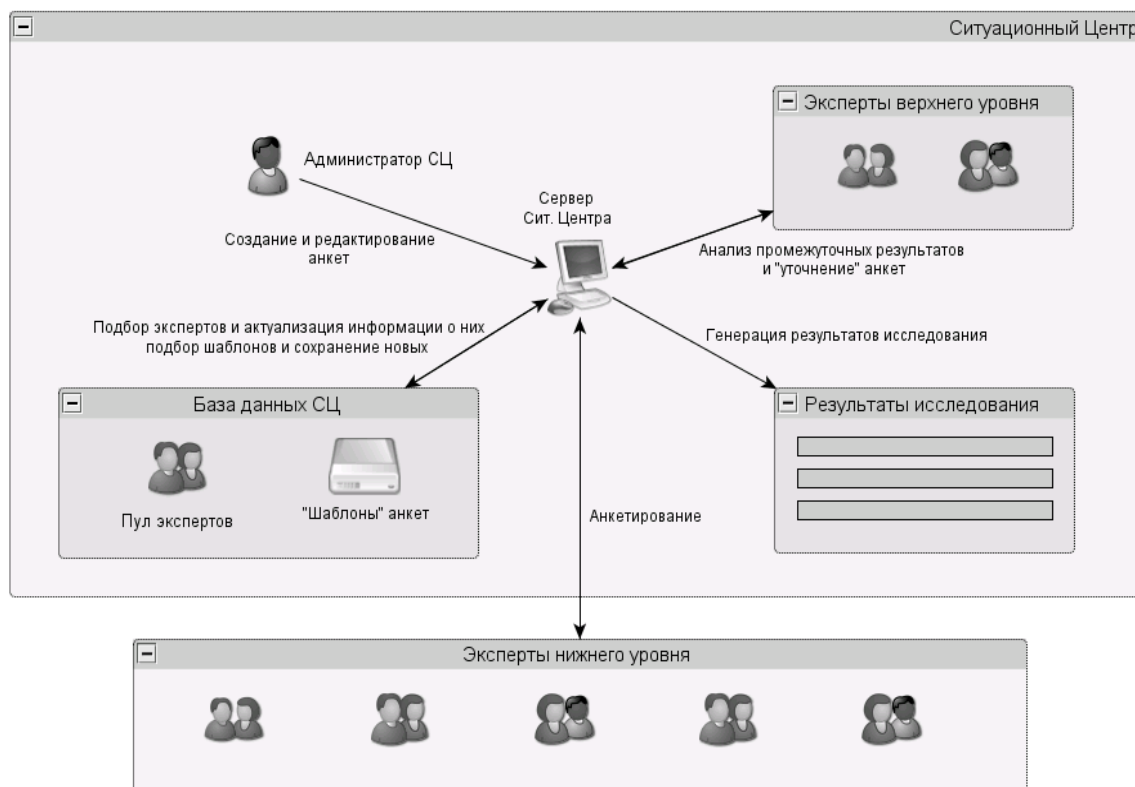


Рис.2 Процесс проведения форсайтнго исследования по методу Дельфи с использованием Ситуационного центра

### Заклучение

Следует отметить, что форсайтные исследования являются не просто видом прогнозирования или предвидения, а скорее способом активного «конструирования» будущего. Поскольку в результате их очерчиваются возможные горизонты, которые могут быть достигнуты лишь при вложении определённых средств и организации систематической работы. Кроме того, Форсайт всегда предусматривает участие многих экспертов, напрямую заинтересованных в решении исследуемых проблем, что позволяет выработать у них видение их конкретных усилий для достижения поставленных в исследовании целей. Ну и, конечно, результатом правильного Форсайт-проекта обычно является не просто прогноз, а выработанные дорожные карты с разработанными практическими мерами по приближению выбранных стратегических ориентиров. Все это делает Форсайт ценным инструментом для разработки стратегии развития не только стран-лидеров, но и развивающихся стран.

В нашей стране методы форсайтных исследований недостаточно развиты на наш взгляд. Причиной тому может быть недостаточное понимание правительством важности разработки оптимальных приоритетов развития и как следствие недостаточное финансирование. Но, кроме того, причиной может быть и недостаточный опыт и следственно культура таких исследований, слишком разновекторные взгляды экспертов на предмет исследования и т.д. Детально проблемы, характерные для Форсайт-исследований, рассмотрены в работе [Малицкий, 2006]], которая представляет результаты опросов 2005-2006 годов по 15 тематическим направлениям, в которых принимали участие более 600 экспертов – одного из самых масштабных Форсайт-проектов в Украине.

На наш взгляд СЦ может быть тем инструментом, который способен помимо автоматизации большинства процедур исследования обеспечить достаточно четкий регламент работы группы экспертов верхнего уровня, а также обеспечить устойчивость и сходимость результатов их работы.

---

### Библиография

---

- [Соколов, 2007] А.В Соколов. Форсайт: взгляд в будущее. «Форсайт», № 1, 2007.
- [Helmer, 1966] O. Helmer. Social Technology. Basic Books. New York, London, 1966.
- [Янч, 1974] Э. Янч. Прогнозирование научно-технического прогресса. «Прогресс», М., 1974.
- [Малицький, 2004] Б.А. Малицький, О.С. Попович, В.П. Соловийов Методичні рекомендації щодо проведення прогнозно аналітичного дослідження в рамках Державної програми прогнозування науково-технологічного та інноваційного розвитку України. «Наукове видання Центру досліджень науково-технічного потенціалу та історії науки НАН України» "Фенікс", К., 2004.
- [Морозов, 2003] А.А. Морозов, В.А. Яценко. Ситуационные центры - основа стратегического управления. «Математические машины и системы», № 1, 2003.
- [Малышев, 2009] О.В. Малышев, С.В. Симонов. Формирование и использование пула экспертов ситуационного центра. «Системи підтримки прийняття рішень. Теорія і практика», К., 2009.
- [Малицький, 2006] Б.А. Малицький, О.С. Попович. Прогноз науково-технологічного та інноваційного розвитку України. «Фенікс», К., 2006.

---

### Authors' Information

---



**Виталий Вишневецкий** – Институт проблем математических систем и машин НАН Украины, старший научный сотрудник, 03187, г. Киев-187, пр. Глушкова, 42;  
e-mail: vit@imm.sp.kiev.ua

Научные интересы: ситуационные центры, обработка и отображение визуальной информации, обработка сигналов, информационные технологии в медицине и биофизике.



**Сергей Симонов** – Институт проблем математических систем и машин НАН Украины, младший научный сотрудник, аспирант, 03187, г. Киев-187, пр. Глушкова, 42;

e-mail: sergiy.simonov@gmail.com

Научные интересы: ситуационные центры, форсайтные исследования, системы поддержки принятия решений.

## ОБОБЩЕННАЯ ОЦЕНКА ОРГАНИЗАЦИОННЫХ СИСТЕМ

Виталий Вишневецкий, Владимир Калмыков, Олег Малышев

**Abstract** The generalized quantitative evaluation of the organizational system using the DOTMLPF method is considered. The method application allows to estimate, having in mind the system requirements, the invested resources efficiency.

**Keywords:** организационная система, предприятие,

**ACM Classification Keywords:** J.1 Administrative Data Processing

**Conference topic:** International Conference KDS 2010 Knowledge - Dialog - Solution September 06 - 09, 2010,

### Введение

Процесс автоматизации управленческой деятельности любой организационной системы, в частности государственной организационной системы или коммерческого предприятия, предполагает существование методики оценки эффективности этой системы на произвольном этапе ее создания или развития. При этом очень важно, чтобы методика оценки эффективности организационной системы охватывала все аспекты ее деятельности - от целевой функции до взаимодействия с окружающей средой. Оценка работы всякой системы должна отображать готовность выполнять те функции, ради которых эта система, или предприятие, создавались. Готовность выполнять ожидаемые от системы функции зависит от наличия ресурсов, необходимых, для выполнения этих функций.

### Методика DOTMLPF

Известная методика DOTMLPF [1] позволяет определить балльную экспертную оценку эффективности выполнения возложенных на организационную систему (предприятие) функций на некотором состоянии ее развития, используя универсальный набор ресурсов. Конкретное содержание ресурсов для каждой системы, безусловно, индивидуально, но всякий набор ресурсов для произвольной системы или предприятия может быть представлен в таком виде (см. табл. 1).

Таблица 1. DOTMLPF.

Литера	Термин	Перевод	Комментарий
1	2	3	4
D	Doctrine	Доктрина	Фундаментальные принципы, нормативные документы согласно которым функционирует система.
O	Organization	Организация	Организационная структура, положенная в основу системы.
T	Training	Обучение (тренинг)	Процессы поддержки существующих и приобретения новых необходимых умений.
M	Materiel	Материальная база	Средства (в т.ч. транспортные), инструменты, приспособления, запасные части, и тому подобное, исключая недвижимость.
L	Leadership	Руководство	Способность руководить.
P	Personnel	Персонал	Сотрудники, необходимые для функционирования системы.
F	Facilities	Недвижимость	Недвижимость, которая принадлежит системе, и используется ею.

Дадим более развернутое описание составляющих схемы DOTMLPF:

1. Доктрина (Doctrine) - фундаментальные принципы, с помощью которых Система руководит своей деятельностью для определенных проектом целей. Доктрина содержит законы и положения на общегосударственном и других уровнях, которые касаются деятельности Системы, официальные постановления ведомств, стандартные методики и процедуры выполнения работ, управляющие директивы, а также разные положения и приказы, относительно организации работ.
2. Организация (Organization) - организационная структура Системы, которая охватывает все уровни Системы, начиная с верхнего - руководящего уровня.
3. Учеба (Training) - Процессы учебы и приобретения/поддержки знаний и навыков персоналом Системы. Существенными являются все аспекты, например, предмет учебы, квалификация инструктора, учебные материалы, оборудование и помещения, предоставленные для этого.
4. Материальная база (Materiel) - все элементы, включая автомобили, специальные средства, соответствующие запасные элементы и части, инструменты, вспомогательное оборудование (но исключая недвижимое имущество), необходимые для того, чтобы оборудовать, управлять и поддерживать подразделения Системы, которые выполняют свои задания.
5. Руководство (Leadership) - способность руководства Системы управлять и влиять на подчиненный персонал.
6. Персонал (Personnel) - сотрудники, необходимые для функционирования Системы.
7. Недвижимость (Facilities) - недвижимость, которая находится на балансе Системы: здания, сооружения, системы помещений, а также земельные участки.

Методика предусматривает возможность декомпозиции требований к системе в процессе ее оценки на составляющие.

Различаются такие варианты получения значения оценки, как получение в результате измерения, или в результате определения экспертом.

В случае определения оценки экспертами для выполнения этой работы обычно привлекается группа экспертов. Понятно, что разные эксперты могут дать разные оценки (при этом каждый эксперт должен аргументировать свою оценку). Окончательная оценка определяется как среднее арифметическое отдельных оценок.

Наличие того или другого ресурса определяется методом экспертных оценок с помощью специальных таблиц. Например, оценка наличия ресурса "Доктрина" определяется по таким правилам (см. табл. 2). Аналогично оцениваются другие виды ресурсов. Все эти ресурсы используются для реализации требований, определенных для системы.

**Таблица 2. Критерии для оценки состояния ресурса "Доктрина"**

Оценка/Общее описание		Критерии
0	Нет	Не существует в письменном виде изложенной доктрины, которая определяет тактику, политику, методы, правила, регламенты и процедуры для данной возможности.
1	Очень низкая	Менее 30 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты и процедуры для данной возможности. Эта официальная доктрина не распространяется устно. Доктрина частично негативно влияет на осуществляемую деятельность.

2	Низкая	Приблизительно 30 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты и процедуры для данной возможности. Эта официальная доктрина сообщается устно в неформальном порядке. Доктрина функционирует.
3	Между низкой и средней	30-50 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты для данной возможности. Процедуры данной возможности разработаны и в минимальной мере согласованы. Официальная доктрина и неформальная доктрина, которая еще существует, сообщается устно в официальном порядке. Доктрина функционирует.
4		
5	Средняя	50-70 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты для данной возможности. Процедуры данной возможности разработаны и согласованы. Следующие дополнения регистрируются и добавляются в официальную доктрину.
6	Между средней и высокой	50-70 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты для данной возможности. Процедуры данной возможности разработаны и согласованы. Доктрина начинает внедрять в практику дополнение для разработки прогнозирования возможных изменений в доктрине. Доктрина функционирует эффективно.
7		
8	Высокая	Больше 70 % доктрины представлены как официально изданный документ, который определяет тактику, политику, методы, правила, регламенты для данной возможности. Процедуры данной возможности разработаны и согласованы. Доктрина функционирует с высокой эффективностью и способна действовать в незапланированных ситуациях.
9	Очень высокая	Сочетание тактики, политики, методов, правил, регламентов и процедур, приводит к созданию исчерпывающих синергетических улучшений эффективности и повышения возможностей управления незапланированными ситуациями.
10	Максимальная	Доктрина представляет собой исчерпывающий и интегрированный набор тактик, политик, методов, правил, регламентов и процедур, что касается данной возможности, и поддерживается в актуальном состоянии.

Проиллюстрируем применение метода DOTMLPF на примере системы, предназначенной для предоставления помощи в случае возникновения экстренной ситуации (ЕС) и соответствующего обращения за помощью путем экстренного вызова (ЭВ) по единственному телефонному номеру 112 (далее - Система-112), которая функционирует в большинстве стран Европейского Союза.

В табл. 3 представлены идентифицированные для системы требования (в соответствии с [2]).



Таблица 3. Требования к Системе-112.

№	Требования	Комментарий
1.	Информированность населения о существовании и правила пользование телефонным номером 112	Эффективность применения Системы-112 в целом существенно зависит от информированности населения
2.	Гарантированный и своевременный ответ оператора	
3.	Бесплатность ЭВ	Определяется доктриной
4.	Автоматическое определение местонахождения Абонента	Эффективность предоставления помощи зависит от определения места ЕС.
5.	Обработка ЭВ оператором Системы-112	Эффективность реакции на ЭВ зависит от качества обработки ЭВ оператором.

В таблице 4 приведенный пример расчета экспертной оценки функционирования системы по воображаемым данным.

Таблица 4. Пример расчета обобщенной экспертной оценки

		D	O	T	M	L	P	F	Суммы:
Требования	1	5	7	3	10	10	10	10	55
	2	1	5	7	5	7	6	7	38
	3	7	10	10	10	10	10	10	67
	4	3	4	1	4	7	2	8	29
	5	3	7	4	8	5	7	9	43
Суммы		19	33	25	37	39	35	44	232

В клетках этой таблицы, которые находятся на пересечении строк - требований и столбиков, - видов ресурсов, занесены экспертные оценки готовности каждого из ресурсов для выполнения каждого из требований, подлежащих оценке. Числа в нижней строке показывают суммарные оценки готовности системы по каждому виду ресурса. Числа в правом столбике показывают суммарные оценки готовности системы по каждому из требований. Оценка готовности системы в целом указана в правой нижней клетке таблицы. Поскольку для требования 1 - Информированность населения ресурсы М, L, P, F не влияют на готовность выполнения этого требования, их оценки приняты максимально возможными -10. Действительно, такие ресурсы, как материальная база (M), руководство (L), персонал (P), недвижимость (F) не могут влиять на информированность населения. В то время как от нормативной базы (D), организации (O), учебы (T), информированность населения зависит. Для выполнения требования 3 - Бесплатность экстренного вызова необходима только законодательная база (D), а другие ресурсы (OTMLPF) на результат не влияют, их оценки выбраны также максимальными -10. Условная оценка готовности системы – 232 может быть использована для сравнения с предыдущими и последующими оценками данной организационной системы, то есть определять динамику ее развития.

---

## Заклучение

Методика DOTMLPF позволяет определить состояние готовности к выполнению запрокированных функций организационной системы или предприятия на каком угодно этапе ее существования - ввод в эксплуатацию, реконструкция или штатное функционирование. Методика позволяет оценить эффективность использования вложенных в развитие системы ресурсов на основе оценки выполнения необходимых возможностей, то есть функций системы или предприятия. Следовательно, методика DOTMLPF может быть использована при выполнении больших проектов, будь то государственная организационная система или коммерческое предприятие, для оценки как текущего состояния, так и завершенных систем.

---

## Библиография

1. DOTMLPF <http://www.pg0.com/dnn/Pg0QsetView/tabid/290/smId/671/ArticleID/13/t/DOTMLPF/>
2. Directive 2002/22/ec of the european parliament and of the council of 7 March 2002 on universal service and users' rights relating to electronic communications networks and services (Universal Service Directive).

---

## Authors' Information



**Vitaly Vishnevsky** - head of division, senior researcher, candidate of engineering sciences, Institute of problems of mathematical machines and systems, prosp. akad. Glushkova 42, 03680, Kiev 187, Ukraine; e-mail: [vit@immsp.kiev.ua](mailto:vit@immsp.kiev.ua)

Major Fields of Scientific Research: Information Technologies, Decision Support Systems



**Vladimir Kalmykov** - senior researcher, candidate of engineering sciences, Institute of problems of mathematical machines and systems, prosp. akad. Glushkova 42, 03680, Kiev 187, Ukraine; e-mail: [vl.kalmykov@gmail.com](mailto:vl.kalmykov@gmail.com), [kvq@immsp.kiev.ua](mailto:kvq@immsp.kiev.ua)

Major Fields of Scientific Research: Image Processing, Information technologies



**Oleg Malyshev** - senior researcher, candidate of engineering sciences, Institute of problems of mathematical machines and systems, prosp. akad. Glushkova 42, 03680, Kiev 187, Ukraine; e-mail: [oleg\\_malyshev@ukr.net](mailto:oleg_malyshev@ukr.net)

Major Fields of Scientific Research: Quality Management, Information technologies

## МОДЕЛИРОВАНИЕ СЛОЖНЫХ ИСКУССТВЕННЫХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ КОНЦЕПЦИИ «ИСКУССТВЕННАЯ ЖИЗНЬ»

Богдан Мыслик

**Аннотация:** В статье рассматривается проблема моделирования сложных искусственных систем. Для моделирования выбрана концепция «искусственной жизни» так как больше всего подходит для решения поставленной задачи. Рассмотрены основные модели искусственной жизни: PolyWorld (ПолиМир), Larry Yaeger; Interaction between learning and Evolution, D. Ackley, M. Littman; «Кузнечик», М. С. Бурцев, Р. В. Гусарев, В. Г. Редько. Обусловив все положительные и отрицательные стороны каждой модели определены основные характеристики модели функционирования производственных предприятий которая учитывает все аспекты моделирования. Выполнена формализация основных составляющих алгоритма моделирования, предложены элементы функциональной нагрузки объектов моделирования.

**Ключевые слова:** искусственная жизнь, нейронные сети, моделирование эволюции, искусственная система, нейронная сеть.

**ACM Classification Keywords:** I.2.0 ARTIFICIAL INTELLIGENCE - General

---

### Введение

С развитием производства, внедрения новых технологий приобретают новое качество задачи моделирования искусственных систем. Актуальной является проблема прогнозирования деятельности таких сложных искусственных систем, как производственные предприятия. Для изучения сложных систем применяется компьютерное моделирование [Самарский, Михайлов, 1997]. Различают аналитическое, имитационное и эволюционное моделирование. При аналитическом моделировании изучаются математические (абстрактные) модели реального объекта в виде алгебраических, дифференциальных и других уравнений, а также предусматривают выполнение однозначной процедуры вычисления, которое приводит к их точному решению. При имитационном моделировании исследуются математические модели в виде алгоритмов, которые имитируют функционирование исследуемой системы путем последовательного выполнения большого количества элементарных операций. При эволюционном моделировании создаются системы, которые являются биологически реалистичнее, чем другие системы [Емельянов, Курейчик, Курейчик, 2003]. К ним относят концепцию «искусственной жизни».

«Искусственная жизнь» (ИЖ) - это не что другое, как попытка выучить именно жизнь, живые системы, их эволюцию с помощью программ созданных человеком [Бурцев, 2007]. Основной проблемой ИЖ является созданием вычислительных систем и моделей, которые действуют на базе биологических и эволюционных принципов.

В работе, проведя обзор моделей с применением концепции ИЖ и отметив особенности каждой, предлагается новый подход к созданию моделей. На основе этого смоделировать эволюцию производственных предприятий с поведением, которое базируется на учете внешних факторов и мотивации.

---

### Обзор научных источников

Для моделирования работы предприятий и выбора модели программирования исследуем основные существующие решения:

- PolyWorld (ПолиМир), Larry Yaeger, 1993;
- Interaction between learning and Evolution, D. Ackley, M. Littman, 1992;
- «Кузнечик», М. С. Бурцев, Р. В. Гусарев, В. Г. Редько, 2002.

**ПолиМир, Ларри Ягер** [Yaeger, 1994]. Модель можно изобразить как большой стол, который ограничен стенками по краям, на котором живут искусственные индивиды. Они двигаются, поглощают энергию и репродуцируются. Также индивиды могут вступать в борьбу с другими индивидами, при этом побежденный индивид превращается в источник энергии для победителя. Ларри Ягер предоставил индивидам интересную особенность - все они владеют цветом и разноцветным зрением. Например, если индивид вступает в борьбу, то его цвет становится красным.

Эти свойства позволяют популяции индивидов эволюционировать, что помогает нам увидеть и проанализировать подобные эволюционные процессы. Фактически модель Л. Ягера показала, что:

- сложное цветное зрение полностью могло сформироваться в процессе эволюции;
- возможно эволюционное формирование трех стратегий.

Что касается первой стратегии, то в ней индивиды двигаются прямолинейно с максимальной скоростью, потребляют все виды энергии на своем пути, и репродуцируются со всеми индивидами на своем пути.

Избрав вторую стратегию, индивид остается на месте, а репродуцируется или вступает в борьбу с другими индивидами лишь при условии их приближения.

Третья стратегия встречается только в некоторых экспериментах, в которых индивиды циркулировали за или против часовой стрелки вдоль барьеров, которые ограничивают стол, и это приводило к определенным преимуществам, поскольку здесь индивиды чаще, чем обычно находили других индивидов, с которыми можно взаимодействовать.

Подводя небольшой итог можно сказать, что Ларри Ягер, с помощью своего ПолиМир реализовал не что другое как естественный отбор индивидов и ничего большего. Именно это имеется в виду под словом «Эволюция» в ПолиМире.

**Interaction between learning and Evolution, D. Ackley, M. Littman** [Ackley, Littman, 1992]. В модели предлагается, чтобы индивиды жили в двумерном мире, разбитом на клетки. В клеточках могут располагаться сами индивиды, деревья, хищники, трава и камни. Соответственно индивиды - это главная «единица» мира, а все другое влияет на них.

Хищники атакуют индивидов, причем сильнее, чем индивиды хищников. Средством для защиты от хищников служит дерево, на котором может влезть индивид, если там нет другого индивида. Но и находиться постоянно на дереве опасно, поскольку деревья со временем погибают, убивая индивида, который сидит на нем.

Поведение индивидов управляется их нейронной сетью. Входами нейронной сети является: видимая картина мира и внутреннее состояние индивидов (количество энергии и здоровья). Индивиды видят мир вокруг себя на расстоянии до 4 клеточек в 4-х направлениях (север, юг, восток, запад). Поведение индивидов определяется выходами их нейронной сети. Мир развивается в дискретном времени. Каждый такт времени нейронная сеть определяет выбор действий индивида. Выбор действий очень простой: выбрать одно из 4-х направлений движения. После выбора действия "локальная судьба" поведение индивида в следующий такт времени будет однозначно определено - она зависит только от того, которое есть в целевой клеточке (клеточке за направлением движения).

Например, индивид может просто переместиться в целевую клеточку (если эта клеточка пуста), съесть в клеточке траву (если она там есть), влезть на дерево (если в целевой клетке есть дерево, и на нем нет индивида), удариться о камень, быть ударенным другим индивидом или хищником и тому подобное.

Нейронная сеть индивида состоит из двух блоков:

- блок поведения, то есть блок, который определяет действия индивида в каждый такт времени;
- блок оценки действия, то есть блок, который формирует цель поведения индивида.

Эти блоки позволяют совместить в данной модели учебу и эволюцию. Блок оценки действия «производит» учебный сигнал для блока поведения, который, в свою очередь, вынуждает индивида действовать соответствующим образом. Сам сигнал может быть или позитивным, если блок оценки

«считает», что команда блока поведения улучшит жизнь индивида, или негативным, если ухудшит жизнь индивида. Таким образом, после разных успехов и неудач индивида, его нейронная сеть корректируется и принимает в тех же действиях «верное» решение. Так и осуществляется учеба.

А в процессе репродукции двух индивидов, их нейроне сети передаются потомкам - так осуществляется эволюция в модели.

Результаты работы Д. Екли и М. Литтмана были достаточно интересны. Они запускали модель по очереди: сначала модель только с учебой, потом модель только с эволюцией и наконец, модель полную - и с учебой, и с эволюцией. Стало понятно, что при взаимодействии учебы и эволюции индивиды не вымирали на протяжении миллиона тактов жизни (а один раз число тактов достигло  $9 \cdot 10^6$ ), тогда как при одиночных моделях индивиды вымирали в 5-7 раз быстрее!

**«Кузнечик», М.С. Бурцев, Р.В. Гусарев и В.Г. Редько.** Наиболее удачно выполнить моделирование жизнедеятельности сложных искусственных систем с учетом, кроме обычных эволюционных процессов, еще и фактора мотивации удалось М.С. Бурцеву, Р.В. Гусареву и В.Г. Редьку в модели под названием «Кузнечик» [Бурцев, Гусарев, Редько, 2001].

Для живых организмов свойственная целенаправленность поведения, то есть стремление достичь определенной цели. Для животных - это самое частое выживание и размножение. В своей работе М. С. Бурцев, Р. В. Гусарев и В. Г. Редько пытаются с помощью компьютерной модели ответить на вопрос: «Как целеустремленное поведение могло возникнуть в процессе эволюции жизни на нашей планете?».

В данном проекте авторами используется такое понятие, как мотивация. Ведь мотивация любого живого существа, которое попадает в ту или другую ситуацию, стимулирует выбор правильного решения. Используя мотивацию, в модели исследуется возможный механизм возникновения целеустремленного поведения в процессе эволюции.

Все действия в модели разворачиваются в одномерной клеточной среде. Время дискретное - то есть в каждый такт времени выполняется одно действие. В клеточках, с определенной достоверностью, растет трава. Есть популяция индивидов, в которых есть потребность в энергии, за счет питания и потребность в размножении. Энергия индивида тратится при выполнении следующих действий, причем на выполнение разных действий тратится разное количество энергии:

- быть в состоянии покоя (отдыхать) - наименьшее количество энергии;
- есть (пополнять свой энергетический ресурс) - в два раза больше энергии;
- двигаться, то есть перемещаться на одну клеточку вправо или влево - еще в два раза больше расходов энергии, чем на питание;
- прыгать через несколько клеточек в случайную сторону - в 5 раз больше расходов, чем при движении;
- репродуцироваться - в 5 раз больше расходов энергии, чем при движении.

Каждая потребность индивида характеризуется количественно мотивацией. Например, если индивид видит рядом другого индивида и его энергетический ресурс достаточен для репродукции, он отмечает себя цветом как готовый к размножению, если второй индивид делает то же - происходит репродукция. В результате появляется новый индивид, который берет часть энергетического ресурса от родителей. Каждый индивид имеет свою нейронную сеть, которая имеет специальные входы от мотиваций.

Нейронная сеть индивида содержит один слой нейронов (Рис. 1). На входы нейронной сети поступают сигналы от сенсорных входов. Входы и нейроны связаны по принципу "все со всеми", то есть каждый нейрон получает сигналы от всех входов. Нейроны формируют исходные сигналы, которые определяют действия индивида.

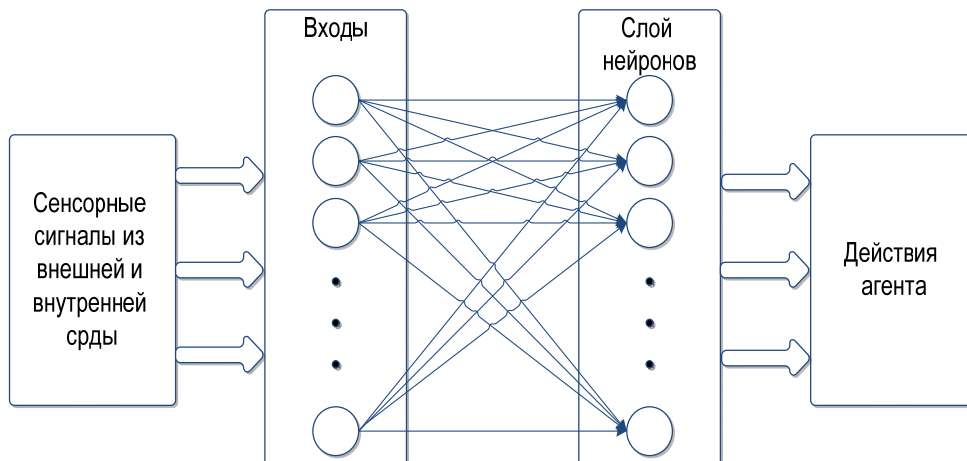


Рис. 1. Структура нейронной сети.

Рассмотрим входные сигналы нейронной сети.

Допускаем, что индивид получает из внешней среды следующую информацию:

- наличие травы во всех клеточках поля зрения;
- есть ли индивид в клеточке слева;
- есть ли индивид в клеточке дело;
- интенсивность синего цвета (что характеризует мотивацию к репродукции) соседних индивидов в ячейке слева и дело.

Кроме того, из внутренней среды индивид получает информацию о его мотивации.

Таким образом, мы имеем  $3+1+1+2+2 = 9$  входных сигналов нейронной сети.

Слой нейронов определяет действия индивида. Каждый нейрон отвечает одному действию. Будем считать, что индивид в данный такт времени выполняет то действие, которое отвечает максимальному исходному сигналу нейрона. Перечислим действия индивида:

- отдых (нахождение в состоянии покоя);
- движение в соседнюю ячейку влево или вправо;
- прыжок (через несколько ячеек);
- питание;
- репродукция с одним 2-х возможных соседей.

Принимая во внимание, что действия 2,5 имеют два варианта, видим, что есть  $1+2+1+1+2 = 7$  разных действий индивида. Соответственно, нейронная сеть индивида содержит 7 нейронов.

Входы и нейроны имеют полные межслоевые синаптические связи, таким образом, есть  $7 \times 9 = 63$  веса.

За счет нейронной сети управляется поведение индивида и так же происходит эволюция индивидов - геном (набор весов нейронной сети) потомка формируется, в результате репродукция, на основе генома родителей с помощью рекомбинации и мутации. При уменьшении ресурса к нулю - индивид умирает. Для анализа были смоделированы два варианта: индивиды с мотивацией и индивиды без мотиваций. Так же были заданы разные параметры  $P$  - достоверность случайного появления травы в каждый такт времени. Конечные результаты показали, что популяция индивидов, которые имеют мотивации намного лучше приспособляется к окружающей среде, причем при среднем количестве еды ( $P=1/200$ ) популяция индивидов с мотивацией «находит» достаточно эффективную стратегию выживания, а популяция без мотиваций вымирает полностью.

Результаты логические и понятные, ведь поведение индивида без мотивации такая: увидев еду - съесть ее, увидев соседа - репродуцироваться с ним, а при отсутствии всего, стоять и ничего не делать (отдыхать), что и приводит индивидов к неминуемой гибели при малом или среднем количестве еды. Когда же у индивида появляется мотивация, в процессе эволюции индивид начинает действовать примерно за такой схемой: мало ресурса - искать еду или отдыхать, много ресурсу - выполнять любые действия. За счет данной схемы популяция выживает намного эффективнее, чем при случае без мотиваций.

Модель «Кузнечик» совмещает в себе и все предыдущие свойства моделей (взаимодействие, учеба и эволюция) и мотивацию, что позволяет наиболее точно и адекватно имитировать деятельность сложных эволюционных систем. Для производственных предприятий это дает возможность экспериментально определить эффективность их деятельности в условиях конкуренции и предварительно заданных параметрах окружающей среды.

### Формализация основных составляющих алгоритма моделирования

Чтобы смоделировать эволюции искусственных систем [Ackley, Littman, 1992], [Бурцев, Гусарев, Редько, 2001], [Букатова, 1991] нужно инициализировать жизненное пространство. Для этого введем условные обозначения.

Существует четыре типа элементов:

- производственные предприятия:  $PV = \{pv_1, pv_2, \dots, pv_{n_1}\}$  во множественное число подвижных элементов, их свойства: источником энергии  $E^M$  являются элементы из множества  $M$ ; источником энергии  $E^R$  являются элементы из множества  $R$ ; источником энергии  $E^D$  являются элементы из множества  $S$ ; они сами являются источником энергии  $E^D$  для элементов из множества  $R$ ; имеют способность к размножению;

- трудовой ресурс:  $R = \{r_1, r_2, \dots, r_{n_2}\}$  во множественное число неподвижных элементов, их свойства: является источником энергии  $E^R$  для элементов из множества  $PV$ ;

- ресурс материалов:  $M = \{m_1, m_2, \dots, m_{n_3}\}$  во множественное число неподвижных элементов, их свойства: является источником энергии  $E^M$  для элементов из множества  $PV$ ;

- потребители:  $S = \{s_1, s_2, \dots, s_{n_4}\}$  грамм множественное число неподвижных элементов, их свойства: является источником энергии  $E^D$  для элементов из множества  $PV$ .

Каждый элемент имеет соответствующие атрибуты значения которых изменяется с течением времени:

1.  $pv_i (e_{pv_i}^t, e_{pv_i}^D, e_{pv_i}^M, e_{pv_i}^R, e_{pv_i}^P - \text{значение энергии}), i = \overline{1, n_1}$ .
2.  $r_i (e_{r_i}^t, e_{r_i}^C), i = \overline{1, n_2}$ .
3.  $m_i (e_{m_i}^t, e_{m_i}^C), i = \overline{1, n_3}$ .
4.  $s_i (e_{s_i}^t, e_{s_i}^P), i = \overline{1, n_4}$ .

Значение энергии  $E^M$  показывает количество материалов, которые находятся на предприятии.  $E^R$  - количество работников.  $E^D$  - количество денежных ресурсов предприятия.  $E^C$  - цена на услуги или материалы.  $E^P$  - цена на продукцию.  $E^T$  - часовая характеристика элемента.

## Моделирование функционирования системы разнотипных элементов

Для моделирования искусственной жизни элементов в жизненном пространстве нужно определить связи между ними и параметры, на которые они влияют (Рис.2). Производственные предприятия, взаимодействуя с элементами ресурсов, пополняют запас материалов для производства товаров, отдавая взамен денежный эквивалент полученным ресурсам. Элементы рабочей силы увеличивают количественную характеристику рабочих на предприятиях, забирая определенное количество денежного ресурса от дохода предприятия. Элементы потребителей забирают товары на предприятиях и возвращают определенное количество денег в зависимости от цены на товар и количества товара.

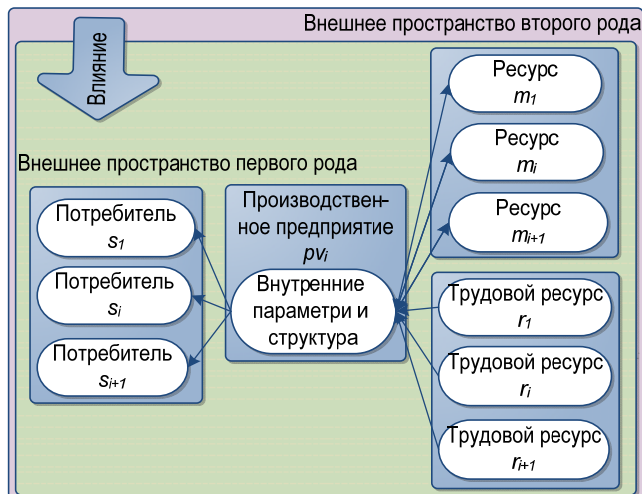


Рис. 2. Схема взаимодействия элементов моделирования

Основные алгоритмические особенности моделирования заключаются в:

1. При взаимодействии элементов из множества  $PV$  с элементами множества  $M$  его значение энергии  $E^M$  увеличивается на значение, что отвечает количеству ресурса за который было заплачено, а значение энергии  $E^D$  уменьшиться на соответствующее количество денег.
2. При взаимодействии элементов из множества  $PV$  с элементами множества  $R$  этот элемент уничтожается, а значение энергии  $E^R$  элемента из множества  $PV$  увеличиться на 1.
3. При взаимодействии элементов из множества  $S$  с элементами множества  $PV$  значение энергии  $E^P$  этого элемента уменьшиться на значение, которое отвечает количеству купленных товаров за какие было заплачено, а значение энергии  $E^D$  увеличивается на соответствующее количество денег.

## Выводы

В данной работе проведен обзор моделей с применением концепции ИЖ. Определены преимущества и недостатки каждой модели и предложено осуществить композицию позитивных факторов. На основе этого представлено моделирование эволюции производственных предприятий, что и было целью исследования.

Выполнена формализация основных составляющих алгоритма моделирования, предложены элементы функциональной нагрузки субъектов моделирования.



В дальнейшем планируется определить параметры влияния внешней и внутренней сред и инициализировать их. Также планируется создать «мозг» предприятия по аналогии модели «кузнечик» в виде нейронной сети, который будет определять действия индивида в зависимости от факторов которые на него влияют (внешние и внутренние).

---

### Литература

---

1. [Самарский, Михайлов, 1997] Самарский А.А., Михайлов А.П. Математическое моделирование: Идеи. Методы. Примеры. – М: Наука, 1997. – С. 320
2. [Емельянов, Курейчик, Курейчик, 2003] Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. – М: Физматлит, 2003. – С. 432
3. [Бурцев, 2007] Бурцев М.С. Пасквиль на эволюцию // Философско-методологические проблемы ИИ. Материалы постоянно действующего теоретического междисциплинарного семинара / Под ред. Е.В. Середкиной. – Пермь: Изд-во ПГТУ, 2007. – С. 129-143
4. [Yaeger, 1994] Yaeger L. Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or Polyworld: Life in a new context. In Langton C.G. (ed). Artificial Life III. – Addison-Wesley, 1994. – P. 263-298.
5. [Ackley, Littman, 1992] Ackley D., Littman M. (1992). Interactions between learning and evolution. // Langton, C. G., Taylor, C., Farmer, J. D., and Rasmussen, S. (Eds.) Artificial Life II. Reading. – MA: Addison-Wesley, 1992. – P. 487-509.
6. [Бурцев, Гусарев, Редько, 2001] Бурцев М.С., Гусарев Р.В., Редько В.Г. Нейросетевая модель эволюционного возникновения целенаправленного адаптивного поведения // Сб. научн. трудов III Всеросс. конф. «Нейроинформатика-2001», часть 1. – М.: МИФИ. – 2001. – С. 153-160.
7. [Букатова, 1991] Букатова И. Л. Эвоинформатика: Теория и практика эволюционного моделирования / И. Л. Букатова, Ю. И. Михасев, А. М. Шаров; АН СССР, Ин-т радиотехники и электрон. – М.: Наука, 1991. – 205 с.

---

### Информация об авторе

---



**Богдан Мысник** – аспирант Черкасского государственного технологического университета, город Черкассы, бульвар Шевченка, 460; e-mail: [Setne@list.ru](mailto:Setne@list.ru)

Отрасли научных исследований: информационные технологии проектирования, моделирования искусственных систем, искусственный интеллект.

## О ЦЕЛОСТНОСТИ ОБРАЗА: ДОФОРМАЛЬНОЕ ИССЛЕДОВАНИЕ

Юрий Валькман

**Abstract:** В работе анализируется свойство целостности образа. В значительной степени работа базируется на исследованиях психологов. Анализируются различные проявления целостности и характеристики образов, которые обеспечивают синтез целостных структур. Показано, что это свойство характерно не только для первичных (перцептивных) образных структур, но для вторичных (образов представления и воображения). Мы полагаем, что без поддержки целостности структур образной информации в вычислительной среде невозможно моделирование процессов образно-понятийного мышления в компьютерных технологиях.

**Keywords:** Образ, мышление, понятие, структура, система, целостность, модель, компьютерная технология, база знаний, вычислительная среда.

**ACM Classification Keywords:** A.1 Information Technology for Knowledge Management; E.1 Process-oriented Knowledge Structuring; H.1.1 Systems and Information Theory; I.2.4 Knowledge Representation Formalisms and Methods; J.4 Social and Behavioral Sciences; M.4 Intelligence Metasynthesis and Knowledge Processing in Intelligent Systems

### 1. Введение

Данная работа является продолжением исследований процессов образного мышления (см., в частности, [Валькман, 2008; Валькман, 2009; Валькман, 2010]). Объектом этих исследований является технологии управления образными знаниями в интеллектуальных системах. Здесь, предмет исследования – свойство целостности образной информации. Цель исследования – разработка моделей представления структур образной информации и методов их использования в практической деятельности.

В последней работе [Валькман, 2010] в центре нашего внимания были структуры образной информации. Структуры образов, явно и неявно, определяются целостностью образной информации, и наоборот.

Целостное, структурное описание объекта (знака) состоит из обобщенных описаний структурных частей объекта и обобщенных пространственных отношений между ними. Обычно разделяют обобщения и укрупнения как что-то независимое, при этом в проблематике искусственного интеллекта (ИИ) чаще рассматриваются обобщения (индукция-дедукция). Укрупнения, т. е. взаимодействия между частями и целым (анализ-синтез), рассматриваются реже.

Рассмотрим некоторые свойства и характеристики целостности структур образной информации.

Поскольку мы исследуем процессы мышления, то логично использовать это понятие в определении психологов [Веккер, 1998]: ОБРАЗ — чувственная форма психического явления, имеющая в идеальном плане пространственную организацию и временную динамику. Такие образы различные исследователи еще называют «мысленными» (imagey, в отличие от pattern), «чувственными», «вторичными», «внутренними». Важно подчеркнуть их значительное отличие от образов в кибернетическом понимании (в проблематике распознавание образов). Фактически, в кибернетике образ определяется как понятие в логике.

В психологии в настоящее время понимание механизма создания различных образов базируется в основном на отнесении их к различным психическим функциям (восприятию, представлению, воображению). Поэтому, психологи выделяют, в частности, три класса образов:

- образ восприятия — отражение в идеальном плане внешнего объекта (сцены), воздействующего на органы чувств,
- образ представления — отражение (вспоминание) объекта без его наличия (сенсорного контакта с ним) и

- *образ воображения* — вымышленный образ, данный в представлении, но не имеющий аналогов в реальной действительности и, поэтому, никогда ранее не воспринимавшийся.

Различие их усматривается обычно в динамике соотношения чувственных и понятийных компонентов, в преобладании единичного или общего, можно говорить и о трех уровнях абстракции (см. [Валькман, 2008]). Схема отношений между этими тремя классами образов представлена на рис. 1.

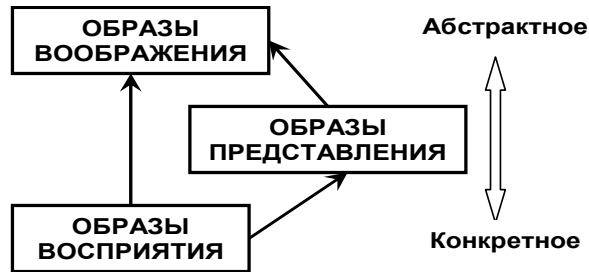


Рис. 1. Схема отношений трех классов образов.

## 2. Отношение «часть - целое»

Понятие *целостности* явно неявно опирается на отношения частей и целого. Вообще, *часть* и *целое* — философские категории, выражающие отношение между некоторой совокупностью предметов и отдельными предметами, образующими эту совокупность.

Категории часть и целое определяются посредством друг друга: *часть* — это элемент некоторого *целого*; *целое* — то, что состоит из *частей*.

Со времен античности проблема соотношения части и целого не только обсуждалась в философии, но в той или иной форме возникала во многих сферах науки и даже общественной жизни. В философии эта проблема формулировалась в самом общем виде: *что является более фундаментальным, исходным, важным — целое или его части?* Решения этой проблемы постепенно оформились в виде двух альтернативных позиций, противоположные исходные принципы которых образовали антиномии целостности. Философская позиция, сводящая целое к его частям и рассматривающая свойства целого только как сумму свойств его частей, получила наименование *меризма* (от греч. *meros* — часть). Противоположная позиция, подчеркивающая несводимость целого к его частям, обретение целым новых свойств по сравнению с его частями, называется *холизмом* (от греч. *holos* — целый). Холистская позиция проявилась в теории эмерджентной эволюции, в витализме, в гештальтпсихологии [Вертегеймер, 1987].

Мереологические (и таксономические) аспекты структур образной информации мы рассмотрели в работе [Валькман, 2010]. Здесь акцентируем внимание на холизме, или гештальте образов, т.е. образ первичен, его компоненты вторичны. Но, конечно, подходы «образ → компоненты» и «компоненты → образ» нельзя противопоставлять и рассматривать изолированно. Они взаимозависимы и дополняют друг друга.

При исследовании целостности образов мы должны «спуститься с высот» философского анализа этой категории и обратить внимание на то, что отношения «часть-целое» может отражать различные аспекты.

Интересный подход к анализу целостности предложен в работе [Лукашевич, 2007] в приложении к лингвистическим структурам. Мы намерены использовать его при построении образных структур. Основная идея авторов состоит в том, чтобы классифицировать различные способы, которыми части соотносятся к своему целому, введением шести различных типов отношений *меронимии*, различаемых на основе признаков *функциональности*, *гомеоморфности* и *отделимости*:

*Функциональные* части ограничены своей функцией, в их пространственном или временном положении. К примеру, ручка чашки может быть расположена в ограниченном числе позиций, если она должна функционировать как ручка.

*Гомеомерные* части представляют собой объект того же типа, что и их целые, к примеру, кусок – пирог, тогда как негомеомерные части отличны от своих целых, например, дерево-лес.

*Отделимые* части могут, в принципе, быть отделенными от целого, к примеру, ручка – чашка, тогда как неотделимые не могут, например, сталь – велосипед.

На основе комбинации этих трех признаков можно выделить следующие типы отношений *часть-целое*:

- *Компонент/Интегральный объект*: Интегральные объекты характеризуются своей структурой, тогда как их компоненты отделимы и имеют специальную функциональность. Например, Колеса – это части автомобиля, фонология – это часть лингвистики.
- *Член/Коллекция*: Члены не играют никакой функциональной роли по отношению к своему целому, они отделимы: дерево – часть леса.
- *Порция/Масса*: Целое рассматривается как гомогенный агрегат и его порции подобны целому: Кусок – это часть пирога.
- *Материал/Объект*: Материал, из которого сделан объект, не отделим от этого объекта, не имеет функциональной роли и негомеомерный: Сталь – часть велосипеда.
- *Фаза/Деятельность*: Фаза, как компонент, имеет функциональную роль, но не отделима: ложка – часть процесса еды, глотание – часть процесса еды.
- *Место/местность*: часть гомеомерна, так как каждая часть пространственного региона подобна целому региону, но не отделима: Оазис – это часть пустыни.

При всем разнообразии подвидов отношения *часть-целое*, существуют классы отношений, которые не рекомендуется путать с отношениями *часть-целое*. Например, к таким отношениям относится отношение «*местонахождение внутри*»: тот факт, что некто находится в комнате, не означает, что этот некто является частью комнаты. Другим видом, отношений, которые необходимо отделять от отношения *часть-целое*, является *отношения присоединения*: тот факт, что электрическая лампа подключена к электрической сети не делает ее частью этой электрической сети.

Заметим, в анализе структур образной информации мы должны рассматривать отношение *часть-целое* еще и на уровне использования различных модальностей: *зрительный канал, слуховой, тактильный и т.д.* Причем, это относится не только к образам восприятия, но и к образам представления и воображения., Интересно также рассмотреть отношение *часть-целое* в структурах: формальные модели, их интерпретация вербальная, графическая и т.д. В образных структурах представляют интерес мультимедийные системы: песни, фильмы и т.п.

Конечно, построение формального аппарата представления образных структур в компьютерных технологиях для моделирования процессов мышления требуют более детальных исследований.

---

### 3. Проблема целостности

---

Целостность - это характеристика любого объекта, всегда состоящего из взаимодействующих частей и обладающего новыми качествами. В системном анализе подчеркивается, новые свойства целого, не сводимые к свойствам частей, каким-то, в общем случае неизвестным, образом определяются свойствами частей и их взаимодействием. Состав, свойства и взаимодействия элементов (частей) каким-то, в общем случае неизвестным, образом определяются свойствами целого. Основным обязательным свойством является *организация* (в данном случае, *образов*), обеспечивающая *устойчивое* существование объекта (системы) во времени. Строящаяся в мозге модель проблемной среды должна содержать целостные представления объектов и ситуаций. Признаковые описания также существуют в модели среды, но эти описания целостными не являются.

Можно ли как-то представить двусторонние взаимодействия между уровнем целого и уровнем частей? Ученые-синергетики исследуя эти взаимодействия, пришли, вообще, к новой парадигме. *Понятие целостности* в этой парадигме приходит на смену понятию *полноты*. В мире живого *полнота*

*недостижима, а мягкость, свойственная целостности, просто необходима.* Без нее невозможно примирить тенденции к суверенности и взаимозависимости частей современного мира, невозможно совместить единство целого и свободу частей [Баранцев, 2009]. Природная тяга к целостности есть тяга к жизни. Стремясь к гармонии целого, человек вызывает недостающие компоненты силой воображения. Заметим, речь идет о создании принципиально новых формальных систем. Может семиотических?

---

#### 4. Структурная целостность

---

Целостность обычно рассматривают с точки зрения ее отношения к частям, при этом стремятся раскрыть неразрывность и взаимообусловленность частей и целого. Структурную целостность образной информации мы рассматривали в [Валькман, 2010]. Ее еще называют в системном анализе – *внутренней*.

Психологи давно заметили (см., например, в [Петровский, 1998]), что образ (его структура и ее построение), как одна из психических реалий, несводим ни к физическим, ни к физиологическим процессам. Уже первые исследования в этом направлении определили разграничение двух существенно различных разрядов психических образов – *сенсорного* и *умственного* (чувственного и мыслимого). И *умственные образы* издавна обозначались непсихологическими терминами – такими, как *понятие* (в логике), *значение слова* (в филологии), *модель* (в формальных науках и технологиях) и т.д.

---

#### 5. Функциональная целостность

---

В этом разделе мы отступаем от традиционного подхода и рассматриваем целостность в ее отношении к внешнему окружению, к среде, т.е. в функциональном аспекте. Такую целостность естественно назвать *функциональной*. В теории систем ее иногда называют *внешней*.

Проблема функциональной целостности занимает существенное место при анализе семиотических систем, в той его части, где рассматриваются вопросы интерпретации знаков и знаковосочетаний (семантика). Значение знака всегда относительно и существует лишь как отношение знака к предмету. С целостными свойствами понятий нашего языка мы сталкиваемся также всякий раз, когда хотим дать им определение. Известно, что *понятие можно определить, лишь подведя его явно или неявно под более широкое понятие*, находящееся на более высоком уровне иерархии понятий. Этот более высокий уровень играет роль своего рода "*лингвистического окружения*", или "*лингвистической среды*", без которой содержание понятий лишается четкости и определенности.

В общей теории систем понятие функциональной целостности с самого начала кладется в основу теории. Оно играет здесь фундаментальную роль наряду с принципом иерархичности. Анализируя понятие системы, В. Н. Садовский рассматривает целостность и иерархичность как равноправные компоненты и ставит их рядом с точки зрения основополагающего значения для теории систем (см. в [Марков, 1982]). Тем самым указывается, что между принципом целостности и принципом иерархичности существует органическая связь.

Иерархическое строение систем в методологическом контексте выступает как следствие функционального характера целостности. С этой точки зрения относительно обособленный объект, рассматриваемый в рамках более широкой системы объект – среда, может трактоваться как уровень иерархии в этой последней системе. Вторым уровнем является окружающая среда. Соответственно этому систему *объект – среда* можно изобразить двумя концентрическими окружностями. Если часть среды, в которой функционирует система (а точнее, ее ближайшее окружение), в свою очередь может быть описана как целостность, то получаем уже *трехуровневую иерархическую структуру*, которую можно изобразить соответственно тремя концентрическими окружностями. И так далее.

Функциональная целостность обуславливает относительную самостоятельность, автономность отдельных подсистем в рамках иерархической структуры. Эта автономность в известном смысле неизбежна, как неизбежно то, что всякий объект, раз он существует, обладает целостными характеристиками, некоторым собственным поведением.

Таким образом, *автономность, целостность, поведенческие характеристики* какого-либо уровня в иерархической системе невозможно понять, изучая структуру только этого уровня, как невозможно понять, что такое стоимость, изучая физико-химические свойства товара. Функции уровня имеют межуровневую природу, выступая как структурные свойства всей иерархической системы, и с этой точки зрения представляют собой основу для проведения структурного анализа системы. Одновременно структура системы может рассматриваться как результат функционального синтеза, т.е. синтеза целостных свойств элементов и уровней системы.

И, таким образом, структуры представления образной информации, онтологий понятийных структур, тезаурусов вербальных структур, вообще любых систем имеют иерархический характер (см. теории и приложения в [Гладун, 1994; Хокинс, 2007; Белов, 2009]). В образных структурах мы можем выделить три слоя в уровнях иерархии; в соответствии с тремя классами образов. Условная схема иерархических отношений образов представлена на рис. 2.

Представляется целесообразным подчеркнуть четыре аспекта в этой структуре.

*Во-первых*, это - иерархия сложности, а не управления, т.е. в ней отражаются отношения «много к многим» между соседними уровнями. Это - *не древовидная структура*. Древовидной она становится тогда, когда рассматривается конкретный образ. Так, например, используется эффективный аппарат растущих пирамидальных сетей [Гладун, 1994]. На древовидные структуры опираются конвергентные и дивергентные процессы в [Белов, 2009].

*Во-вторых* (как следствие, во-первых), число компонент на различных уровнях иерархии не уменьшается по мере перемещения снизу вверх, а наоборот, *растет*. Это обусловлено, в частности, тем, что порождение новых образов может производиться с использованием различных комбинаций базовых («нижележащих») признаков (свойств).

*В-третьих*, уровни иерархии определяются целостностями, входящими в рассматриваемый, конкретный образ. Поэтому, в целом, иерархическая структура – нерегулярна и гетерогенна.

*В-четвертых*, человек не всегда в состоянии четко определить, о каком классе образов идет речь в конкретных случаях. Поэтому, границы в слоях носят *условный, нечеткий характер*; они «размыты».

Следующие два раздела, в значительной степени, базируются на материале фундаментальной монографии [Веккер, 1998].

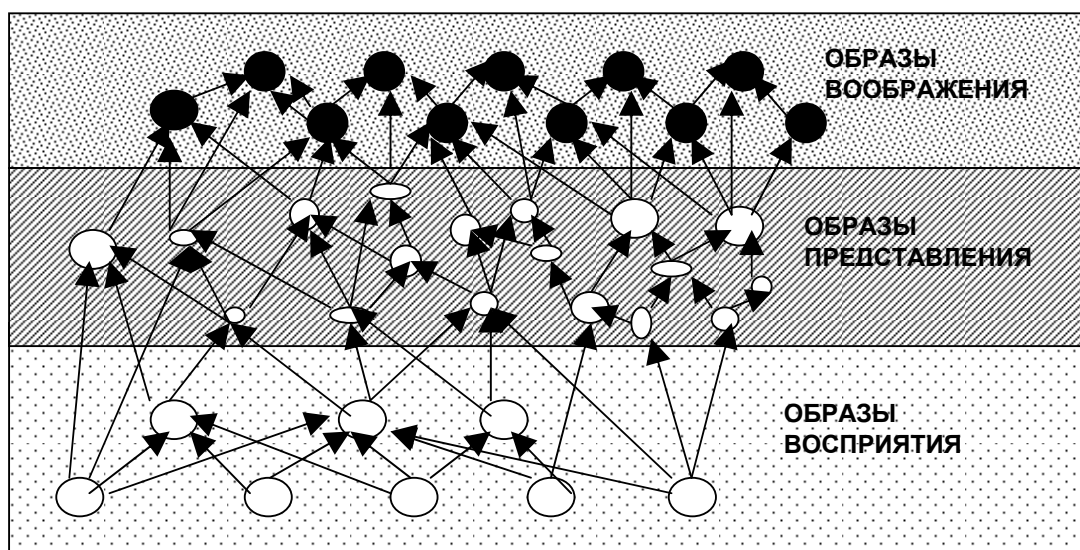


Рис. 2. Условная схема иерархической структуры образной информации.

## 6. Целостность образов восприятия

Целостность восприятия — свойство восприятия, состоящее в том, что всякий объект, а тем более пространственная предметная ситуация воспринимаются как устойчивое системное целое, даже если некоторые части этого целого в данный момент не могут быть наблюдаемы (например, тыльная часть вещи) [Веккер, 1998]. Проблема *целостности восприятия* впервые была экспериментально исследована представителями гештальтпсихологии [Вертегеймер, 1987]. Однако здесь целостность восприятия выступала как изначальное его свойство, определяемое законами сознания. Российская психология рассматривает целостность восприятия как *отражение целостности, объективно присущей воспринимаемому* [Якиманская, 1985]. Образ, формирующийся в процессе отражения действительности, обладает высокой избыточностью. Это значит, что некоторая совокупность компонентов образа содержит информацию не только о самой себе, но и о других компонентах, а также об образе в целом. Так, наблюдатель, который по условиям восприятия мог наблюдать голову и плечи прохожего, воспринимает положение его рук, туловища и даже характер его походки. А это уже образы воображения.

*Восприятием называется отражение в сознании человека предметов и явлений окружающего мира при их непосредственном воздействии на рецепторы в форме целостных образов.*

Мы воспринимаем игровые действия футболистов, которые в борьбе за мяч осуществляют задуманную ими тактическую комбинацию. Альпинист воспринимает предметы и пространственные отношения между ними, когда смотрит на открывшуюся его взору горную долину, видит находящиеся вблизи деревья и скалы, текущую несколько дальше горную речку и далекие остроконечные пики гор на горизонте. Студент воспринимает речь преподавателя, читающего лекцию.

Эти примеры показывают, что в процессе восприятия мы получаем образы воспринимаемых вещей и явлений окружающего мира. *Образы восприятия* отличаются следующими особенностями [Веккер, 1998].

1. Объективация, вынесение образа вовне. Воспринимая какой-нибудь предмет (дерево, книгу и т. д.), мы осознаем его не как наше субъективное психическое переживание, а как объективный, вне нас существующий предмет.

2. Целостность. Восприятие всегда имеет целостный характер: в получаемых в процессе восприятия конкретных образах предметов внешние явления отражаются в их целостности, в органической совокупности их свойств и качеств.

Образ восприятия не механическая сумма частей или элементов, из которых состоит воспринимаемый предмет, а образ самого объекта во всей его целостности. С самого начала, с первого момента восприятия мы имеем дело сразу же, мгновенно с образом целостной вещи, а не создаем его путем суммирования элементов. Наоборот, расчленение образа на элементы — вторичный процесс, дополняющий целостное восприятие и идущий вслед за ним. Сначала мы видим дом, а затем уже различаем этажи и другие части строения. Сначала мы слышим мелодию во всей ее целостности, а затем уже выделяем составляющие ее аккорды и музыкальные тоны.

Целостный характер восприятия обусловлен природенной нашему мозгу способностью видеть в воспринимаемом предмете то, что составляет его особенность как *целого предмета*, а затем уже выделять присущие ему элементы.

Во многих случаях конкретный характер частей и элементов *целостного предмета* не имеет для его восприятия существенного значения и может быть легко заменен другими конкретными особенностями без того, чтобы была потеряна целостность восприятия. Так, мы воспринимаем мелодию как целостно одну и ту же, несмотря на то, что она сыграна на различных музыкальных инструментах или в различных регистрах, что, как известно, полностью меняет конкретный характер составляющих ее элементов. Необходимо лишь, чтобы при исполнении сохранялось обуславливающее *целостность мелодии* определенное соотношение составляющих ее музыкальных тонов. Любую букву алфавита мы всегда узнаем как таковую, несмотря на значительное расхождение в транскрипции. Необходимо лишь, чтобы при всех этих расхождениях сохранялось характеризующее целостность предмета соотношение частей.

*Целостность образа* всегда обусловлена тем, что мы непосредственно воспринимаем в предмете обуславливающие эту целостность особенности, хотя не всегда можем определить, в чем они состоят. Так, мы сразу же и легко различаем на фотографии женское лицо от мужского и в то же время с трудом можем сказать, чем мы при этом руководствуемся.

Благодаря целостному характеру восприятия мы всегда восполняем в своем восприятии отсутствующие элементы и части воспринимаемого предмета.

3. *Подвижность*, отсутствие длительной фиксации определенных частей, невозможность на сколько-нибудь длительное время сохранить постоянство образа, который все время находится в движении, изменении. Поскольку процесс восприятия всегда совершается во времени, образ воспринимаемого предмета отличается подвижностью, изменчивостью; это не застывшее, статическое изображение, но всегда меняющееся в своих характерных чертах. Например, при восприятии дерева в сознании человека в каждый данный момент отражается преимущественно то та, то другая часть предмета: сейчас в восприятии дерева наиболее отчетливо выступает его своеобразный ствол; через секунду в том же образе дерева более ярко отразится его крона.

4. *Константность*. При всей своей подвижности и изменчивости образы воспринимаемых нами предметов отличаются определенным постоянством (константностью), несмотря на значительную вариативность условий, в которых протекает процесс восприятия. Так, лист писчей бумаги мы воспринимаем всегда как белый, хотя его окраска может принимать различные оттенки в связи с изменением условий освещения. Стол воспринимается нами обязательно как имеющий квадратную или прямоугольную верхнюю часть, хотя в данный момент мы можем видеть его под таким углом зрения, когда его верхняя поверхность от нас скрыта и т.д.

5. *Осмысленность*. Образы восприятия всегда имеют четко выраженное определенное смысловое значение («я вижу дерево, море, человека» и т. д.). Мы всегда относим наблюдаемый предмет или явление к определенной группе или классу предметов, а не отображаем их в своем восприятии как нечто изолированное, не имеющее отношения к другим явлениям. Осмысленность восприятия достигается прежде всего тем, что конкретный образ воспринимаемого явления мы тотчас же обозначаем *словами*.

Хотя гештальт-психология [Вертгеймер, 1987] в своих теоретических позициях исходила из изначальности и первичности целого по отношению к элементам, в ее экспериментальном материале – в полном соответствии с несомненным общенаучным положением о том, что целое состоит из элементов и поэтому в каких-то своих характеристиках неизбежно от них зависит – содержатся факты, воплощающие второй аспект этого соотношения – зависимость способов группировки элементов в целое от характеристик самих элементов. Эксперименты выявляют несколько факторов или эмпирических законов такой группировки элементов в целостную структуру [Веккер, 1998]:

- *Фактор близости*. При прочих равных условиях в целостную структуру объединяются элементы по признаку наименьшего расстояния между ними.
- *Фактор замкнутости*. В единую перцептивную структуру объединяются элементы, в совокупности составляющие замкнутый контур или замкнутую трехмерную поверхность.
- *Фактор хорошей формы*. Объединению подвергаются элементы, образующие в целом особый предпочтительный класс так называемых хороших форм, таких, например, как круг или прямая линия, т.е. тел или фигур, обладающих свойством симметричности, периодичности, ритма и т.д.
- *Фактор коллективного движения*. К объединению в группу тяготеют элементы, совершающие совместное перемещение (стая птиц, эскадрилья самолетов и т.д.).
- *Фактор однородности*, заключающийся в том, что детерминантой объединения элементов оказываются их общие пространственные или модальные характеристики; в единую группировку входят компоненты одной формы, одного цвета и т.д.



Таким образом, все выделенные факторы, на основе которых элементы связываются в *целостную структуру*, суть не что иное, как разные частные формы общего объединяющего их начала – *однородности по какому-либо из пространственно-временных или модальных признаков*.

Все эти факторы и свойства перцептивных образов, хорошо известные психологам, приведены здесь для определения направлений исследования процессов синтеза и анализа образных структур с целью их моделирования в компьютерных технологиях.

Обратим внимание, что практически все свойства и факторы базируются на использовании долговременной памяти («обработанной, систематизированной и структурированной информации» - образах представления). Поэтому, строго говоря, эти свойства нельзя относить к свойствам образов восприятия. Это – свойства взаимодействия перцептивных образов и образов представления.

---

## 7. Целостность образов представления

---

Представления – это необходимое посредствующее звено, смыкающее первосигнальные психические процессы, организованные в форму образов различных видов, и второсигнальные мыслительные или рече-мыслительные психические процессы, составляющие уже "специально человеческий" уровень психической информации [Веккер, 1998]. Уже рассмотрение такого важнейшего свойства первичных образов, как обобщенность, которая не случайно завершает перечень эмпирических характеристик перцепта и является "сквозным" параметром всех психических процессов, привело к вопросу о необходимой взаимосвязи восприятия и памяти.

Поскольку *обобщенность образа* выражает отнесенность отображаемого в нем объекта к определенному классу, а класс не может быть содержанием актуального, т.е. в данный момент совершающегося, отражения, обязательным посредствующим звеном здесь является включенность *апперцепции*, т.е. образов, сформированных в прошлом опыте и воплощенных в тех извлекаемых из памяти эталонах, с которыми сличается каждый актуальный перцепт [Веккер, 1998]. Такие эталоны и есть *образы представления*, аккумулирующие в себе признаки различных единичных образов. На основе этих признаков строится "*портрет класса объектов*", и тем самым обеспечивается возможность перехода от перцептивно-образного к понятийно-логическому отображению структуры класса предметов, однородных по какой-либо совокупности своих признаков. Заметим, именно на реализации свойства *апперцепции* и *акцепторах* Анохина [Анохин, 1980] базируется модель мозга «*память - предсказание*» Д. Хоккинса. Но это было самостоятельное открытие (см. «синюю чашку» в [Хоккинс, 2007]).

Исследование образов представления сталкивается с существенными трудностями как в исходном пункте анализа – при описании их основных эмпирических характеристик, так и на этапе теоретического поиска закономерностей. Л. М. Веккер предложил трехуровневую классификацию характеристик образов представления. Он их называет вторичными образами. Здесь приведен фрагмент этой классификации, который интересен, с нашей точки зрения, для моделирования процессов образного мышления.

### 1. Особенности пространственно-временной структуры вторичных образов.

1.1. *Пространственная панорамность* заключается в том, что целостное воспроизведение пространственной структуры объекта во вторичном образе не ограничивается объемом перцептивного поля и выходит за его пределы.

Так, пространственный массив, охватываемый единым топографическим представлением, превосходит по угловым размерам объем перцептивного поля, а представление об отдельном объекте может охватывать те компоненты или стороны последнего, которые при непосредственном восприятии находятся за пределами поля зрения.

1.2. В отличие от перцептивного образа, существенной особенностью которого является выделение фигуры из фона, не допускающее, однако, их взаимного отделения, в представлении *фигура может не соотноситься с определенной координатой пространственного фона*, а фон может быть отделен от фигуры ("пустое пространство").

1.3. *Выпадение абсолютных величин* проявляется в двух моментах:

- в не сохранении числа однородных элементов (например, числа колонн в представлении об Исаакиевском или Казанском соборе);
- в нарушении воспроизведения абсолютных размеров отображаемого пространственного массива и в особенности размеров отдельного объекта.

1.4. Преобразование геометрической формы в топологическую схему во вторичном образе имеет *разнообразные проявления*. Оно выражается в схематизации образа и др. свойствах (см. [Веккер, 1998]).

1.5. *Симультанность, или "временная панорамность"*, представлений заключается в том, что компоненты временной и двигательной последовательности имеют тенденцию преобразовываться во вторичном образе в одновременную структуру, в которой эта последовательная динамика очень затуманена или не воспроизводится совсем. По отношению к слуховым музыкальным представлениям, которые воспроизводят не последовательное развертывание, а *одновременноцелостную структуру* музыкального произведения.

1.6. Совершающиеся во вторичном образе сдвиги в воспроизведении длительности установлены в многочисленных исследованиях, данные которых обобщены в закономерность *"чем более заполненным и, значит, расчлененным на маленькие интервалы является отрезок времени, тем более длительным он представляется"*.

2. Особенности модальных характеристик вторичного образа состоят в том, что во вторичном образе цвета происходит перестройка, аналогичная сдвигам восприятия цвета в затрудненных условиях: *образ смещается в сторону основных цветов спектра, а отдельные конкретные оттенки из образа выпадают тем в большей мере, чем более длительным является срок хранения образа*.

Остальные три характеристики образов представления Л. М. Веккер называет производными.

3. Первой из вторичных характеристик представления является *неустойчивость*. Будучи по своей сути (как проявление непостоянства) отрицательным эквивалентом, или выражением дефицита константности, свойственной перцептивному образу, неустойчивость представления, хорошо известная каждому по собственному опыту, заключается в колеблемости и текучести его компонентов.

3. *Фрагментарность представлений*, состоит в том, что "при внимательном анализе или попытке установить все стороны или черты предмета, образ которого дан в представлении, обычно оказывается, что некоторые стороны, черты или части вообще не представлены". Сопоставление с эмпирическими характеристиками образов восприятия, легко увидеть, что если *неустойчивость представления есть аналог неполной константности*, то *фрагментарность представляет собой эквивалент неполной целостности* или выражение ее дефицита в представлении по сравнению с восприятием.

4. *Параметр обобщенности (интегрированности)*, будучи общей характеристикой не только всех видов образов, но и вообще всех психических процессов, имеет во вторичных образах свою отчетливо выраженную специфичность по сравнению с обобщенностью первичных. Если образ восприятия, какова бы ни была степень его генерализованности, всегда является обобщенным изображением того конкретного единичного объекта, который воздействует на анализатор, то вторичный образ, в силу того, что представляемый объект не воздействует на органы чувств, *может быть не только единичным, но и общим*. Это означает, что, воплощая в себе целый ряд ступеней обобщенности образа, на высших из этих ступеней представление освобождается от "прикованности" к единичному объекту и "может быть обобщенным образом не единичного предмета или лица, а целого класса или категории аналогичных предметов" [Веккер, 1998].

Производный характер второй подгруппы характеристик по отношению к первичным эмпирическим характеристикам представления выражается в том, что *неустойчивость, фрагментарность и обобщенность*, как и их перцептивные гомологи – константность, целостность и обобщенность, охватывают все три первичных параметра вторичного образа: пространственно-временную структуру,

модальность и интенсивность. Мы полагаем, что процессы моделирования образного мышления должны включать процедуры выявления неустойчивости и фрагментарности образов представления и обобщенность

---

## 8. Целостность образов воображения

---

С нашей точки зрения целостность образов воображения принципиально отличается от целостности других классов образов. Хорошим примером такой целостности может служить стихотворение. Стихотворение в целом создает поэтический образ, которого не может создать в отдельности ни одно из входящих в него слов. Связи между словами в стихотворении определяются стихотворным размером и правилами грамматики. Контекст стихотворения, поэтический образ, создаваемый им, определяют смысловое содержание входящих в него слов и часто способны совершенно изменить его по сравнению с обычным употреблением. Как раз такого рода целостности имел в виду холизм и, поэтому, совершенно справедливо подчеркивал несводимость целого к его частям. И, если целостность первых классов образов, обусловлена целостностью объектов внешнего мира, то целостность образов этого класса всегда создается сначала в нашей голове. Например, известно, что композиторы или писатели способны видеть или слышать все произведение сразу, а математики - видеть сразу всю структуру доказательства. И только потом писать партитуры, строить доказательства, т.е. декомпозировать эту целостность в алгоритм. Согласно легенде, великий Гаусс говорил: «Решение у меня есть уже давно, но я еще не знаю, как к нему прийти». Известно множество гипотез, которые «ждут превращения» в теоремы. Выигрышный ход в шахматах – это тоже целостность! Видимо, все творчество – это построение целостных структур в воображении. Здесь целесообразно вспомнить о работе [Саймон, 1972].

---

## 9. Заключение

---

Все отмеченные свойства образных представлений чрезвычайно плохо и неадекватно моделируются в символично-логической парадигме - и в логических теориях, и в компьютерных моделях. Эти теории и модели пригодны только для работы с внешними представлениями. Образные, внутренние представления, которые нам интересны и специфика работы, с которыми составляет суть эффективности мозга, видимо, необходимо изучать на моделях нейронного типа.

Основное отличие базы знаний моделирования образного мышления от традиционных приложений интеллектуальных технологий заключается в необходимости поддержки высокой динамичности структур знаний в компьютерной среде. Поэтому к процедурной компоненте модели образных представлений предъявляются особые требования. Смена фокуса внимания, расширение, сужение, перемена контекста интерпретации, поддержка процессов смысловой индукции требуют соответствующей реструктуризации системы образов. Таким образом, известные аппараты нейросетей не подходит для поддержки этих структур. В процессе выполнения операций мышления появляются новые структуры (т.е. новые нейроны и новые отношения).

Общеизвестно, моделирование мышления в вычислительной среде эквивалентно реализации искусственного интеллекта. Данное исследование предпринято для построения концептуальных моделей хранения и использования образных структур в решении проблем. Однако, здесь изложены только некоторые принципы и концепции. В последующих публикациях будет представлен структурно-содержательный анализ образа на уровне доформального исследования. Далее, на основе этих моделей, будут построены формальные структуры, которые являются базой для создания компьютерных технологий моделирования образно-понятийного мышления и рассуждений. При этом, мы согласны с автором [Шамис, 2006] в том, что *«мышление – это не изощренные алгоритмы решения задач, а специфический принцип их реализации в живом мозге»*. Поэтому, речь идет о моделировании процессов образного мышления при решении некоторых проблем и задач.

---

## Литература

---

- [Анохин, 1980] Анохин П.К. Узловые вопросы теории функциональных систем Москва: Наука, 1980,
- [Баранцев, 2009] Баранцев Р.Г. Комплексирование целостности // Глобализация: синергетический подход. Москва: РАГС, 2002;
- [Белов, 2009] Белов Ю. А., Рабинович З. Л. Память человека и мышление - образное и символическое. Концептуальное модельное представление. // Доклады Академии Наук, Москва: Наука, 2009, том 427, № 6.
- [Брунер, 1997] Брунер Дж. Психология познания. – Москва: Прогресс, 1977.
- [Валькман, 2008] Валькман Ю. Р. Анализ понятия образ: отношения «образы – понятия» // Сб. тр. XI Междунар. конференции «Искусственный интеллект (КИИ-2008)», Россия, Дубна, 2008.
- [Валькман, 2009] Валькман Ю.Р. О моделировании образного мышления: классификация, отношения и структуры образов. // Труды IX-й Междунар. конференции "Интеллектуальный анализ информации ИАИ-2009", Киев, 2009.
- [Валькман, 2010] Валькман Ю.Р. О структуре образа: доформальное исследование // Труды X-й Международ. конференции "Интеллектуальный анализ информации ИАИ-2009", Киев, 2010.
- [Веккер, 1998] Веккер Л.М. Психика и реальность: единая теория психических процессов. М.: Изд-во "Смысл", 1998.
- [Вертгеймер, 1987] Вертгеймер М. Продуктивное мышление. – Москва: Прогресс, 1987.
- [Гладун, 1994] Гладун В.П. Процессы формирования новых знаний - София: СД "Педагог 6", - 1994.
- [Лукашевич, 2007] Лукашевич Н.В., Моделирование отношения ЧАСТЬ-ЦЕЛОЕ в лингвистических и онтологических ресурсах. // Информационные технологии. - 2007. - N 12.
- [Марков, 1982] Марков Ю. Г. Функциональный подход в современном научном познании. Новосибирск: Изд-во "Наука", 1982.
- [Петровский, 1998] Петровский А. В., Ярошевский М. Г. Основы теоретической психологии. М.: ИНФРА-М, 1998.
- [Ричардсон, 2006] Ричардсон Дж. Т.Э. Мысленные образы: Когнитивный подход. Москва: Когито – Центр, 2006.
- [Саймон, 1972] Саймон Г. Науки об искусственном. - Москва: Мир, 1972.
- [Хокинс, 2007] Хокинс Д., Блейкли С. Об интеллекте. Москва: Изд. Дом. Вильямс, 2007.
- [Шамис, 2006] Шамис А.Л. Пути моделирования мышления. Активные синергические нейронные сети, мышление и творчество, формальные модели поведения и "распознавания с пониманием". Москва: КомКнига, 2006.
- [Якиманская, 1985] Якиманская И. С. Основные направления исследований образного мышления. // Вопросы психологии, 1985, № 5.

---

## Информация об авторе

---



**Юрий Валькман** – Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, 03680 ГСП, г. Киев, проспект акад. Глушкова, 40; e-mail: [yur@valkman.kiev.ua](mailto:yur@valkman.kiev.ua)

**Область научных интересов:** Моделирование образного мышления, Методы и средства искусственного интеллекта, Интеллектуальные технологии и системы, Базы знаний, Системной анализ, Технологии управления знаниями.

## К ИНТЕГРАЦИИ ОНТОЛОГИЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

**Александр Палагин, Андрей Михайлюк, Виталий Величко, Николай Петренко**

**Аннотация:** Рассматривается проблема системной интеграции онтологий предметных областей с точки зрения формирования общего интегрированного пространства трансдисциплинарных знаний. Разработана методика и алгоритм указанной интеграции.

**Ключевые слова:** онтология предметной области, системная интеграция.

**ACM Classification Keywords:** I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods.

---

### Введение

Новым шагом развития междисциплинарных научных исследований, и теории баз знаний в частности, должно стать теоретически обоснованное объединение (или системная интеграция) уже разработанных как общедоступных онтологий, так и коммерческих баз знаний для разнообразных прикладных задач, проблем, целых предметных областей и трансдисциплинарных знаний общего характера. В [Палагин 2009] сущность системной интеграции сформулирована следующим образом: «Устойчивые знания совокупности научных дисциплин можно представить в форме интегрированной иерархической сети научных теорий (разного уровня развития, содержательности и охвата действительности), составляющие которых, возможно, связаны посредством общих объектов действительности». Там же говорится о цели междисциплинарных исследований – приближение к построению общенаучной картины мира – а также о системной интеграции знаний (онтологий) как одной из важных задач в достижении указанной цели.

Системная интеграция двух и более онтологий представляется далеко не тривиальной задачей. Её алгоритмы не проработаны, требуют тщательной проверки и апробации на представительном множестве онтологий.

Формально онтология представляется тройкой  $\langle X, R, F \rangle$ , где  $X$  – множество концептов,  $R$  – множество отношений между концептами,  $F$  – функции интерпретации концептов из множества  $X$  и отношений из  $R$ . Даная модель носит общий характер, в то время, как на практике пользуются более точными моделями. Одной из таковых является модель Protégé Frames [Noy 2000 (1)], разработанная для использования в редакторе онтологий Protégé [Sachs 2006] и базирующаяся на фреймовом представлении и протоколе ОКВС (Open Knowledge Base Connectivity) [Chaudhri 1998]. Известны и другие применяемые на практике модели, активно используемые в онтологическом инжиниринге, особое место среди которых занимает язык описания онтологических знаний OWL [Dean 2004], разработанный для онтологизации веб-пространства и принятый в качестве стандарта консорциумом W3C. Однако, все современные практически-ориентированные модели эквивалентны модели  $\langle X, R, F \rangle$  или являются ее частными (ограниченными) представлениями, а, следовательно, могут быть представлены в ней, например, согласно схеме соотнесения общепринятой модели и фреймовой модели Protégé (рис. 1). Исходя из этого, для описания операций над онтологиями можно использовать любую из моделей.

В данной работе будет использована фреймовая модель Protégé, так как она имеет высокую степень абстрактности, достаточно распространена и описана во множестве публикаций и документации разработчиков [Protégé project и др.], имеет формальные средства описания в популярном языке OWL [Knublauch 2004], и, кроме того, редактор, основанный на этой модели, является общедоступным как в исходных кодах, так и в бинарном виде для большинства современных платформ [Download Protégé], что дает возможность применять его на практике.

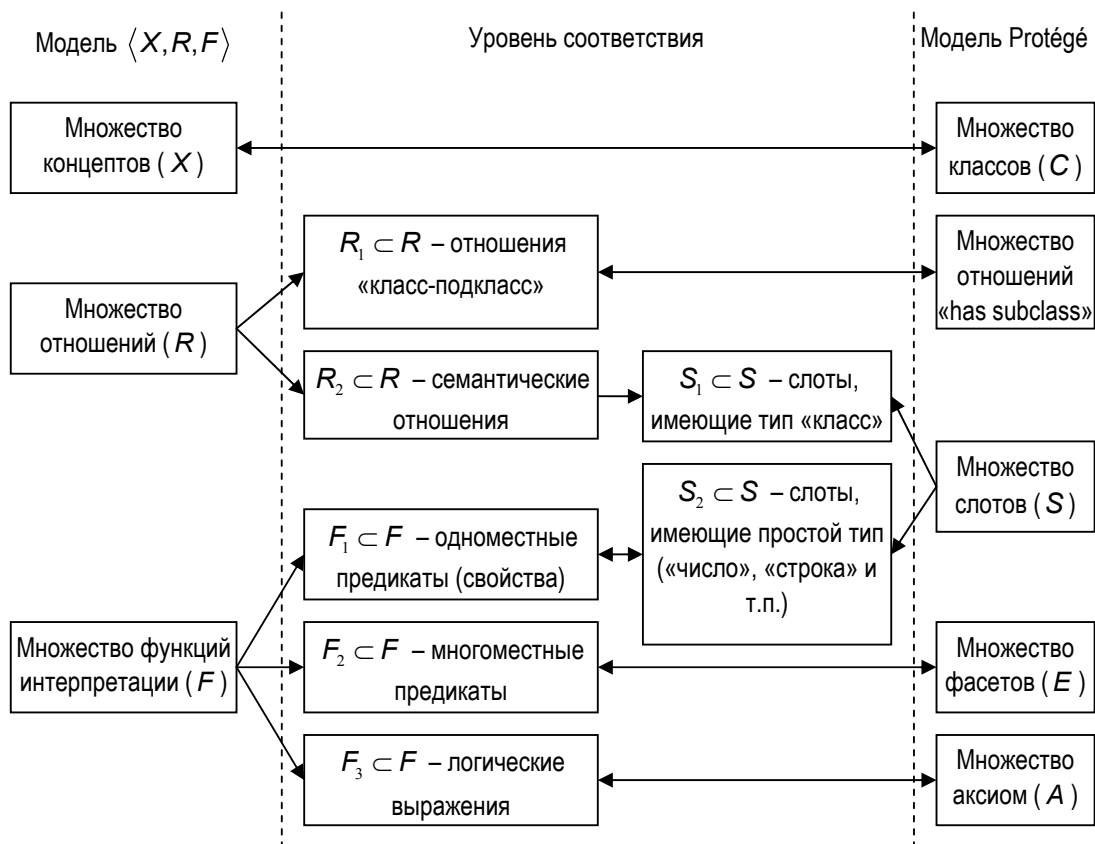


Рис. 1. Схема соотнесения модели  $\langle X, R, F \rangle$  и фреймовой модели Protégé.

### Постановка задачи

Каждая предметная область (ПдО) может содержать подмножество классов, эквивалентных классам понятий других ПдО. Это могут быть как классы, обозначающие общенаучные понятия, так и классы, обозначающие понятия, специфические для каждой ПдО. На основе отождествления некоторых классов (понятий) разных ПдО возможно построение их общей онтологии. На сегодняшний день известно множество онторедкторов, большинство из которых реализуют процедуру построения общей онтологии из двух исходных [Овдей 2004]. Практическое применение такого объединения очевидно – от добавления в онтологию нового, ранее не разработанного фрагмента ПдО, полученного от других разработчиков, до построения прототипа общего интегрированного пространства трансдисциплинарных знаний.

Ряд задач все еще остается открытым, а их решение, как правило, предоставляется пользователю системы (инженеру по знаниям). Кроме того, большинство решений, принимаемых указанными системами относительно эквивалентности классов, являются очевидными, тогда как более сложные случаи подобия пользователь вынужден определять вручную. Целью данной работы является расширение процедур обработки онтологий, качественное улучшение известных программных продуктов манипулирования онтологиями путем повышения уровня автоматизации их работы и адекватности предлагаемых пользователю действий. Для этого разрабатывается, обосновывается и описывается алгоритм построения результирующей онтологии из двух исходных на основе результатов сравнения классов. Кроме того, предлагается обобщение задачи объединения онтологий до процедуры интеграции онтологий, описывающих различные ПдО одного и того же домена предметных областей.

---

### Общий подход к объединению онтологий

---

Из известных операций над онтологиями [Овдей 2004] объединение онтологий является одной из наиболее важных и часто используемых пользователями процедурой. Суть указанной процедуры состоит в построении результирующей онтологии из двух исходных с сохранением первичных онтологических знаний в таком виде, чтобы результирующая онтология была максимально связанной (содержала все отношения между классами) и не содержала дублирующих данных (например, эквивалентных классов).

В наилучшем варианте, алгоритм объединения онтологий ориентирован на автоматическое выполнение необходимых действий без какого-либо участия пользователя. По окончании его работы результат предьявляется пользователю системы в любой системе визуализации (возможно, не связанной с системой объединения). С другой стороны, наиболее популярные на сегодня приложения объединения онтологий [Овдей 2004] предлагают полуавтоматический режим. Он заключается в интерактивном объединении пользователем пар классов, принадлежащих объединяемым онтологиям, объединяя таким образом сами онтологии «поэлементно». Основная роль системы при этом заключается в автоматическом выполнении операций, связанных с объединением классов (создание новых «объединяющих» классов, копирование их слотов, установление родовидовой зависимости, переименование классов и слотов и т.д.). Кроме того, в некоторых подсистемах (плагилах или модулях онторедкторов) объединения онтологий [Noy 2003, McGuinness 2000] помимо выполнения операций, связанных с объединением онтологий, пользователю предлагается набор возможных на данном этапе операций («объединить классы», «скопировать слот» и т.п.).

Алгоритм объединения онтологий можно разделить на три относительно независимых этапа, на каждом из которых учитывается полученный на предыдущем этапе результат:

1. Установление степени эквивалентности между классами двух онтологий. На данном этапе производится сравнение всех классов одной онтологии со всеми классами другой. Качество процедуры сравнения существенно влияет на качество объединения онтологий.
2. Поэлементное объединение множеств классов и множеств слотов исходных онтологий. Основная часть, по-сути, и являющаяся объединением, основывается на степени эквивалентности, вычисленной на предыдущем этапе для каждой пары классов. Этот этап, как правило, выполняется полностью автоматически, так как не требует принятия решений относительно семантики классов и полностью формализован. Выбор алгоритма работы данного этапа объединения онтологий влияет на степень интеллектуальности всей процедуры. Под интеллектуальностью процедуры объединения онтологий понимается возможная степень сложности результирующих графов – родовидовых иерархий (классы, соединенные отношениями «класс-подкласс») и семантических сетей (классы, соединенные семантическими отношениями), которые могут быть построены при объединении, возможную степень их отличия от соответствующих графов исходных онтологий.
3. Проверка результата объединения на корректность. Под корректностью понимается непротиворечивость результирующей онтологии и степень ее соответствия ожиданиям пользователя. Последнее может быть полностью реализовано только самим пользователем, тогда как критерий противоречивости может быть определен формально. Возможные типы противоречий определяются моделью описания онтологии. В простом варианте – это противоречия, связанные с уникальностью имен классов и слотов [Noy 2000 (2)]. В общем случае, противоречия определяются ложностью аксиом и функций интерпретации. Поскольку большинство онторедкторов имеют ограниченные возможности задания логических зависимостей между классами и их отношениями, противоречия таких типов не проверяются в связи с отсутствием в модели функций интерпретации в явном виде (модель Protégé, например, сводит задание функций интерпретации к заданию фасетов).

## Сравнение классов

Для реализации первого этапа процедуры объединения онтологий необходимо разработать операцию сравнения классов двух онтологий. Входными параметрами для операции выступают исходные онтологии ( $O$  и  $O'$ ) и пара сравниваемых классов ( $C \in O$  и  $C' \in O'$ ). Результатом указанной операции является логическое значение («истина» или «ложь»), соответствующее ответу на вопрос «Эквивалентен ли класс  $C$  классу  $C'$  ?». Исходя из сложности внутренней структуры класса и его отношений с другими классами, отметим нетривиальность задачи сравнения классов. Прежде всего, следует определить, на основании чего можно утверждать, что некоторый класс эквивалентен или подобен другому классу. Поэтому, первой частью разработки операции сравнения классов является выбор критериев сравнения.

Наиболее простым способом сравнения классов является *сравнение по имени*. Данный подход является основным критерием большинства существующих модулей объединения онтологий [Овдей 2004]. В наиболее развитых из них (таких как PROMPT) существует возможность нечеткого сравнения имен классов и сравнения с учетом синонимов [Ehrig 2005], однако в случае сложных синонимичных и омонимичных конструкций сравнение классов по данному критерию дает неверные результаты. Причиной таких ошибок является предположение о роли класса в онтологии по его имени, тогда как имя не определяет класс (как термин не определяет понятия), а служит удобной для человека меткой класса.

Более сложным вариантом сравнения классов является сравнение по содержанию (составу), однако во многих случаях отождествление классов (понятий) на основе поэлементного равенства всех их экземпляров является ошибочным, описание причин и примеры чего представлены в [Войшвилло 2007]. Несмотря на это, сравнение объемов классов активно используется в современных системах как вспомогательная информация для установления подобия между ними [Noy 2000 (2), Noy 2001].

Согласно [Войшвилло 2007, Ивлев 2001], понятие полностью определяется своим *содержанием*, что следует из самого определения содержания понятия, а поскольку класс в модели Protégé аналогичен понятию, то это же должно быть справедливо и для классов. Исходя из этого, можно заключить, что при сравнении классов наиболее важной составляющей в их структуре является его содержание, т.е. множество слотов, доменом которых является данный класс. Однако, отношения между множествами (в данном случае, множествами слотов) не ограничиваются отношениями эквивалентности и различия – последнее может представлять включение одного множества в другое, включение второго множества в первое, непустое пересечение двух множеств и отсутствие общих элементов (полное различие). Относительно множеств слотов данных двух классов указанные отношения можно трактовать следующим образом:

1. Совпадение всех слотов класса  $C_1$  со всеми слотами класса  $C_2$  (поэлементное равенство множеств  $\{S\}_1$  и  $\{S\}_2$ ) означает равенство содержания двух классов. Это в свою очередь, свидетельствует об *эквивалентности* самих классов. Назовем данный результат «эквивалентность».
2. Включение множества  $\{S\}_1$  слотов класса  $C_1$  во множество  $\{S\}_2$  слотов класса  $C_2$  ( $\{S\}_1 \subset \{S\}_2$ ) свидетельствует об обобщении содержания класса  $C_2$  до содержания класса  $C_1$ . Это означает, что сам класс  $C_1$  является более общим по отношению к классу  $C_2$ . Исходя из принципа обратной зависимости содержания и объема понятий (увеличение содержания влечет уменьшение объема и наоборот [Войшвилло 2007, Ивлев 2001]), можно заключить, что класс  $C_1$  является одним из *надклассов* класса  $C_2$ . Назовем данный результат «обобщение».
3. Включение множества  $\{S\}_2$  слотов класса  $C_2$  во множество  $\{S\}_1$  слотов класса  $C_1$  ( $\{S\}_1 \supset \{S\}_2$ ) является случаем, обратным к случаю (2), и означает, что класс  $C_1$  является *подклассом*  $C_2$ , т.е. понятие, представляемое классом  $C_1$ , является уточнением понятия, представляемого классом  $C_2$ . Назовем данный результат «уточнение».



4. Пересечение множеств слотов классов  $C_1$  и  $C_2$  ( $\{S\}_1 \cap \{S\}_2 \neq \emptyset$ ) свидетельствует о наличии общих слотов. Это означает, что существует некоторый класс  $C$ , являющийся надклассом для классов  $C_1$  и  $C_2$ , а сами классы принадлежат одному уровню иерархии. Назовем данный результат «частичная эквивалентность».

5. Пустое пересечение множеств слотов классов  $C_1$  и  $C_2$  ( $\{S\}_1 \cap \{S\}_2 = \emptyset$ ) свидетельствует о том, что у первого общего для  $C_1$  и  $C_2$  надкласса отсутствуют слоты. Таким надклассом может быть корень родовидового дерева («Универсум» или класс «THING» в Protégé) или классы категориального уровня (такие как «абстрактное», «материальное», «процесс», «сущность» и т.д. [Палагин 2006]). В этом случае, такие классы можно считать полностью различными. Назовем данный результат «различие».

Описанный механизм сравнения определяет взаимное родовидовое отношение двух классов, что является обобщением отношения эквивалентности и вместо двух вариантов результатов сравнения классов предполагает пять. Однако, такой механизм не применим к классам с пустым множеством слотов – в этом случае (согласно описанному механизму) такие классы должны быть эквивалентны, что, однако, в большинстве случаев неверно. Пустое множество слотов класса говорит о том, что понятие, представляемое классом, не требует описания его содержания в явном виде, либо что его содержание неизвестно. Последний случай следует отнести к некорректности входных данных, так как алгоритм ориентирован на объединение полных онтологий. Случай же с необязательностью указания содержания некоторого понятия возникает при уверенности инженера по знаниям, составляющего онтологию, что оно (понятие) известно интерпретатору онтологии (системе обработки онтологий или другому инженеру по знаниям). В таком случае понятие, а следовательно и представляющий его класс, должны однозначно идентифицироваться своим именем, поэтому предлагается *классы без слотов сравнивать по именам*.

---

### Объединение и интеграция онтологий

---

Основываясь на результатах поэлементного сравнения классов, система может перейти к непосредственному объединению онтологий. Базовый механизм объединения, используемый в современных системах обработки онтологии [Овдей 2004], заключается в поиске классов, эквивалентных, по мнению системы или решению пользователя. Эквивалентные классы добавляются в результирующую онтологию, а их содержание представляет собой множественно-логическое объединение слотов и фасетов исходных классов. При необходимости слоты могут переименовываться, фасеты проверяться на совместимость, и т.д. Именно часть корректного переноса содержания исходных классов в результирующий считается основной задачей процедуры объединения. Несомненно, наличие такой процедуры значительно упрощает процесс получения новой онтологии из двух исходных, по-сравнению с «ручным» режимом. Однако, такой подход носит во многом механический характер «склеивания» отдельных классов без попытки интеллектуализировать этот процесс.

В данной работе предлагается расширить возможности процедуры объединения онтологий за счет обобщения операций, предлагаемых пользователю в существующих системах. Этап перенесения слотов и фасетов исходных классов в соответствующие классы результирующей онтологии остается неизменной, а в большинстве случаев даже упрощается, поэтому детально не будет описан в данной работе (см. [Ной 2000 (2)]). Итак, как было отмечено выше, обобщением отношения эквивалентности классов является отношение родовидовой зависимости: вместо взаимного позиционирования двух классов как эквивалентных или различных предлагается рассматривать эти классы как пару эквивалентных, пару частное-общее, общее-частное, пару классов с общей частью или же пару полностью различных классов (без общей части). Очевидно, пять вариантов взаимного отношения классов как результата операции их сравнения приводят к пяти различным операциям над классами на этапе объединения онтологий:

- если классы эквивалентны, то они представляют одно и то же понятие в онтологии, следовательно, должны быть «склеены» в один;
- если класс одной онтологии является обобщением соответствующего класса другой онтологии, такие классы должны представляться как класс и подкласс соответственно, причем совпадающие слоты должны быть удалены из подкласса, так как они будут унаследованы от надкласса (т.к. отношение "класс-подкласс" является отношением частичного порядка);
- если класс одной онтологии является уточнением соответствующего класса другой онтологии, такие классы должны представляться как подкласс и класс соответственно, причем совпадающие слоты должны быть удалены из подкласса, так как они будут унаследованы от надкласса;
- если классы двух онтологий частично эквивалентны, то они представляют собой схожие понятия, то есть должны иметь общий надкласс, являющийся их обобщением (заметим, этот надкласс не присутствовал ни в одной из исходных онтологий), при этом совпадающие слоты должны быть удалены из подкласса, так как они будут унаследованы от обобщающего класса;
- если классы различны, они должны быть скопированы в результирующую онтологию.

Как было отмечено выше, при объединении применяется итеративный подход, когда после выполнения операции над текущей парой классов система проводит повторное сравнение необработанных классов, исходя из результатов выполненной операции, и предлагает (или сразу выполняет) наиболее подходящую на этом этапе операцию из описанных выше. Поскольку классы исходных онтологий либо полностью копируются в результирующую, либо переносятся в нее с изменением родовидовых связей и удалением общих слотов, то и сравнение классов следует выполнять в рамках результирующей онтологии. В таком случае, возможно построение новой онтологии на базе одной из исходных путем введения в нее классов из второй исходной онтологии.

Описанная процедура качественно отличается от существующих процедур *объединения* – вместо объединения онтологий «склеиванием» эквивалентных вершин происходит взаимное «проникновение» онтологий со значительным изменением исходной структуры, но полным сохранением исходных онтологических знаний. Поэтому данную процедуру более точно было бы называть *интеграцией* (а не объединением) онтологий. Основой процедуры интеграции онтологии в таком случае является не «склеивание» эквивалентных вершин, а интеграция каждого класса интегрируемой онтологии в базовую.

**Алгоритм процедуры интеграции онтологий.** Алгоритм процедуры интеграции онтологий удобно разбить на три части: общая часть, интеграция родовидовых иерархий (интеграция с учетом отношений класс-подкласс) и интеграция классов (интеграция с учетом слотов как элементов содержания классов). Представим блок-схемы этих трех частей процедуры интеграции онтологий (согласно [ГОСТ 19.701-90] и [ISO 5807:1985]) и опишем их алгоритмы.

Общая часть процедуры выполняет подготовительную и завершающую части интеграции. Кроме того, она вызывает процедуру поэлементной нисходящей интеграции (интеграция иерархий). Поэтому назовем эту часть процедуры общим названием «Интеграция онтологий» и представим блок-схему ее алгоритма на рис. 2 (а).

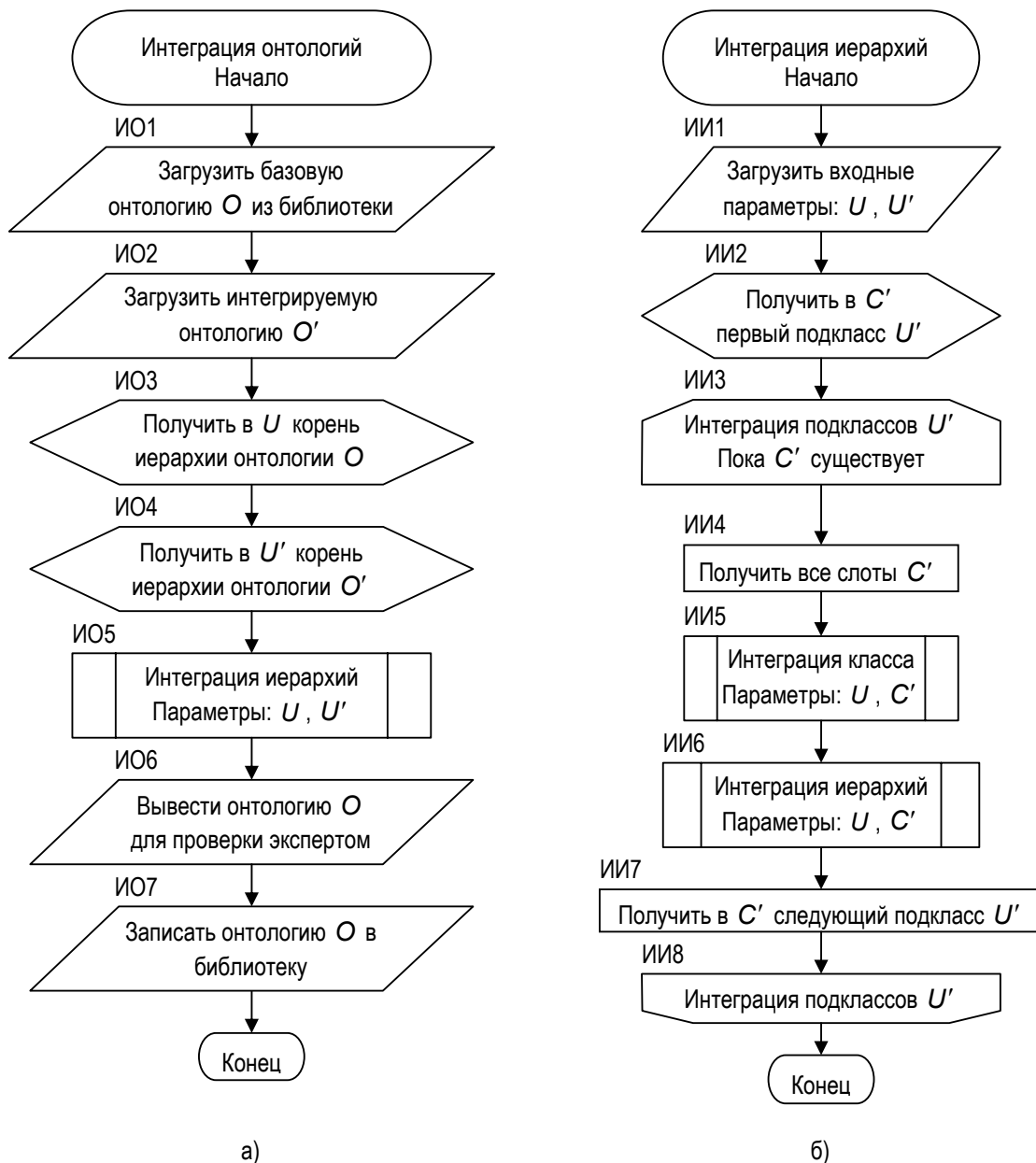


Рис 2. Блок-схема процедур: (а) «Интеграция онтологий», (б) «Интеграция иерархий»

**ИО1.** Выполняется загрузка базовой онтологии  $O$  из библиотеки онтологий ПдО в модуль системной интеграции онтологий.

**ИО2.** Выполняется загрузка онтологии  $O'$ . Источником может быть web-страница с RDF разметкой, носитель данных с сохраненной ранее онтологией, другая система обработки онтологий (например, система поиска документов и составления их онтологий).

**ИО3.** Выполняется инициализация корневой вершины  $U$  базовой онтологии  $O$ .

**ИО4.** Выполняется инициализация корневой вершины  $U'$  интегрируемой онтологии  $O'$ .

**ИО5.** Вызов и передача управления подпрограмме «Интеграция иерархий» для корневых классов базовой и интегрируемой онтологии.

**ИО6.** Процедура «Интеграция онтологий» передает результат работы подпрограммы «Интеграция иерархий» в подсистему визуализации (например, Protégé) для просмотра и проверки инженером по знаниям правильности выполненной интеграции.

**ИО7.** Выполняется запись результирующей онтологии в соответствующую библиотеку.

Рассмотрим алгоритм интеграции иерархий, представленный на рис.2б.

Интеграция структур классов исходных онтологий представляется как интеграция родовидовой иерархии, начиная с корня иерархического дерева. Эту задачу, в свою очередь, удобно выполнять рекурсивно: интегрировать корень иерархии (входящий класс) в базовую иерархию (также заданную своим корнем – входным параметром) и поддеревья каждого из подклассов данного класса вызовом этой же процедуры с текущим подклассом данного класса в качестве вершины иерархии. Для возможности рекурсивного обращения данная часть алгоритма системной интеграции выделена в отдельную процедуру.

**ИИ1.** Процедура «Интеграция иерархий» загружает входные параметры – классы  $U$ ,  $U'$ . Класс  $U$  задает вершину родовидового дерева базовой онтологии, в которую будет интегрироваться иерархия, заданная корневым классом  $U'$ . Сам класс  $U'$  не интегрируется в иерархию  $U$ , так как представляет собой прототип класса  $U$  в интегрируемой онтологии.

**ИИ2.** Задание временной переменной  $C'$  (указателя на класс – итератора для прохождения по списку непосредственных подклассов  $U'$ ) начального значения – первого непосредственного подкласса класса  $U'$  в списке подклассов класса  $U'$ .

**ИИ3.** Начало цикла с предусловием. В цикле осуществляется интеграция текущего подкласса  $C'$  класса  $U'$  в иерархию, заданную классом  $U$ , и всей иерархии, заданной подклассом  $C'$ . Цикл выполняется, пока не будет достигнут конец списка подклассов класса  $U'$ , то есть подкласс  $C'$  как элемент списка «удачно» извлечен (предполагается, что при достижении конца списка операция извлечения следующего его элемента вернет несуществующий «нулевой» элемент).

**ИИ4.** В список слотов класса  $C'$  добавляются слоты всех его надклассов, реализуя, таким образом, принцип наследования. Данный блок необходим, так как подпрограмма «Интеграция класса» интегрирует класс с полным списком слотов в заданную иерархию.

**ИИ5.** Класс  $C'$  интегрируется в иерархию базовой онтологии с вершиной  $U$  вызовом процедуры «Интеграция класса».

**ИИ6.** Рекурсивный вызов процедуры «Интеграция иерархий» относительно класса  $U$  и текущего подкласса  $C'$  для интеграции поддерева интегрируемой онтологии, заданного вершиной  $C'$ , в иерархию базовой онтологии, заданной классом  $U$ .

**ИИ7.** Итерирование по списку подклассов класса  $U'$  – получение следующего элемента списка подклассов класса  $U'$  и запись его в переменную  $C'$ .

**ИИ8.** Переход на начало цикла (блок ИИ3) «Интеграция подклассов» с текущим значением  $C'$ .

Интеграция классов исходных онтологий заключается в размещении данного класса интегрируемой онтологии в иерархию базовой онтологии. При этом учитываются варианты взаимного отношения классов, описанные в подпункте «Базовый алгоритм сравнения классов», и применяются операции, описанные в начале данного пункта. Общая схема работы алгоритма состоит во вставке интегрируемого класса в базовую иерархию на максимально низкий уровень. Проход вниз по иерархии реализован рекурсивно.

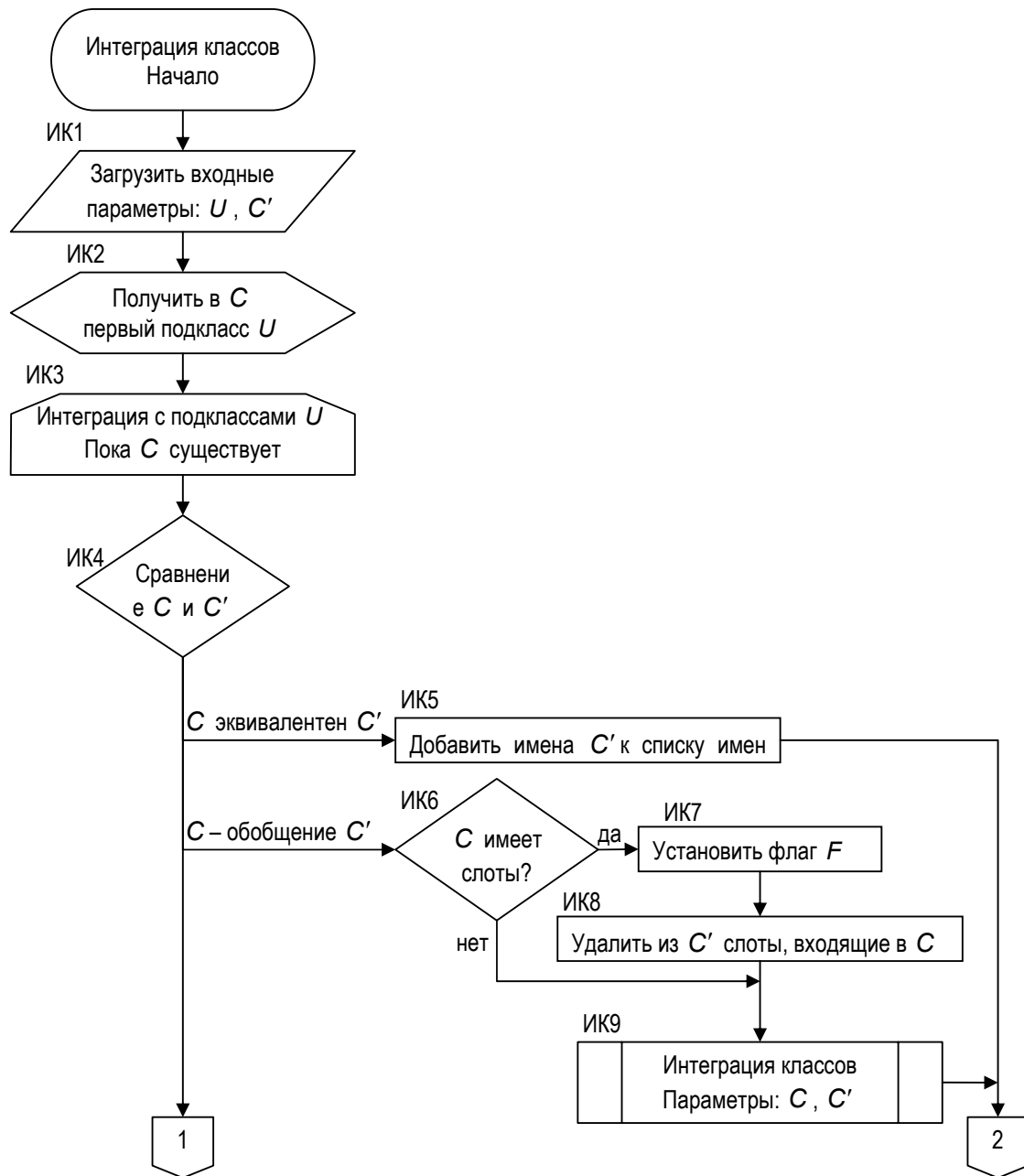


Рис. 2.в. Интеграция классов (начало)

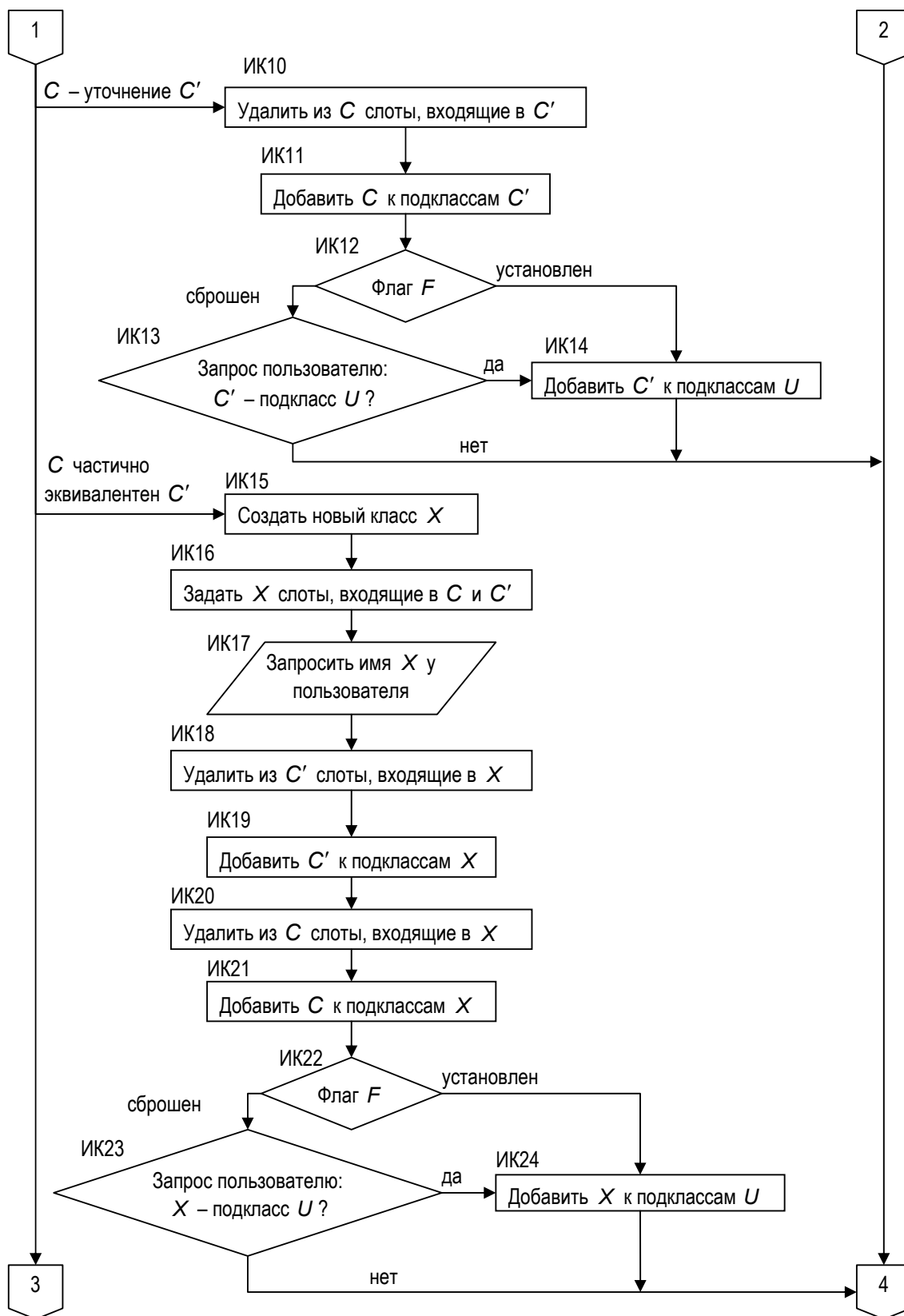


Рис. 2.в. Интеграция классов (продолжение)

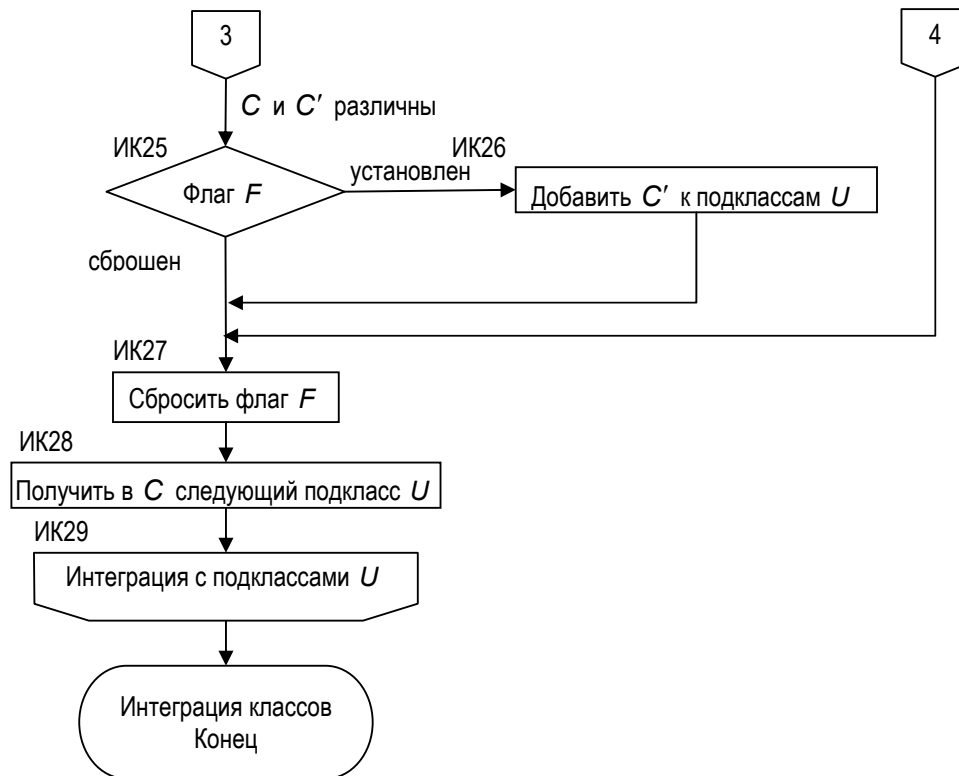


Рис. 2.в. Интеграция классов (окончание)

**ИК1.** Процедура «Интеграция классов» загружает входные параметры – классы  $U$ ,  $C'$ . Класс  $U$  задает вершину родовидового дерева базовой онтологии, в которую будет интегрироваться класс  $C'$ .

**ИК2.** Задание временной переменной  $C$  (итератора для прохождения по списку непосредственных подклассов  $U$ ) начального значения – первого непосредственного подкласса класса  $U$  в списке подклассов класса  $U$ .

**ИК3.** Начало цикла с предусловием. В цикле осуществляется сравнение и взаимное размещение классов  $C$  и  $C'$ . Цикл выполняется, пока не будет достигнут конец списка подклассов класса  $U$ , то есть подкласс  $C$  как элемент списка «удачно» извлечен (предполагается, что при достижении конца списка операция извлечения следующего его элемента вернет несуществующий «нулевой» элемент).

**ИК4.** Сравнение классов  $C$  и  $C'$  согласно механизму, описанному в подпункте «Базовый алгоритм сравнения классов».

**ИК5.** Если классы  $C$  и  $C'$  эквивалентны, к списку имен класса  $C$  добавляются имена  $C'$ . Иными словами, данный блок реализует объединение синонимов. Далее алгоритм переходит к завершающей части цикла (блоку **ИК27**).

**ИК6.** Если  $C$  является обобщением  $C'$ , класс  $C'$  должен находиться ниже по иерархии относительно класса  $C$ . Однако, это может быть неверно, если  $C$  является «возможным обобщением»  $C'$ , то есть принадлежит категориальному уровню и не содержит слотов. Осуществляется проверка на наличие слотов у класса  $C$ .

**ИК7.** Если  $C$  имеет слоты, значит  $C'$  должен быть добавлен в данную ветвь иерархии. Для учета этого при дальнейшем сравнении устанавливается флаг  $F$  (глобальная переменная логического типа).

**ИК8.** Из множества слотов класса  $C'$  удаляются слоты, входящие также в класс  $C$ , поскольку  $C'$  будет добавлен в подклассы класса  $C$  (непосредственно или через другие классы) на следующих этапах работы алгоритма, а значит, унаследует данные слоты.

**ИК9.** Независимо от того, содержит ли класс  $C$  слоты, необходимо сравнить класс  $C'$  с подклассами  $C$ . Для этого рекурсивно вызывается процедура «Интеграция классов» относительно  $C$  и  $C'$ , реализуя тем самым «спуск» класса  $C'$  вниз по иерархии, заданной входящим классом  $U$ . Далее алгоритм переходит к завершающей части цикла (блок **ИК27**).

**ИК10.** Если класс  $C$  является уточнением класса  $C'$ , класс  $C'$  должен находиться выше по иерархии относительно  $C$ . Из множества слотов класса  $C$  удаляются слоты, входящие также в класс  $C'$ , поскольку класс  $C$  унаследует их от класса  $C'$ .

**ИК11.** Класс  $C$  добавляется в список подклассов класса  $C'$ .

**ИК12.** Проверка флага  $F$ , показывающего, была ли установлена в блоках **ИК6-ИК8** необходимость добавления класса  $C'$  в иерархию класса  $U$  (флаг установлен), или «спуск» класса  $C'$  производился через категориальные классы, которые не обязательно являются надклассами  $C'$  (флаг сброшен).

**ИК13.** Если флаг  $F$  сброшен, система не может принять однозначного решения относительно добавления класса  $C'$  в иерархию класса  $U$ : нет никаких признаков, что  $C'$  входит в иерархию  $C$ , но нет и никаких признаков, что  $C'$  не входит в иерархию  $U$ . Данную неоднозначность предлагается решить пользователю, для чего формируется вопрос: «Является ли  $C'$  подклассом  $U$ ?». Если ответ пользователя отрицательный, алгоритм переходит к завершающей части цикла (блок **ИК27**).

**ИК14.** Если флаг  $F$  установлен, либо пользователь указал, что  $C'$  является подклассом  $U$ , класс  $C'$  добавляется в список подклассов класса  $U$ . Далее алгоритм переходит к завершающей части цикла (блок **ИК27**).

**ИК15.** Если  $C$  частично эквивалентен  $C'$ , необходимо создать для классов  $C$  и  $C'$  общий надкласс, содержащий в себе слоты, входящие в оба сравниваемых класса. Создается класс  $X$ .

**ИК16.** В новый класс  $X$  копируются все слоты, общие для классов  $C$  и  $C'$ .

**ИК17.** Поскольку  $X$  не входил ни в одну из исходных онтологий, он не имеет имени. Система запрашивает имя класса  $X$  у пользователя, поясняя, что  $X$  – это обобщение классов  $C$  и  $C'$ .

**ИК18.** Из класса  $C'$  удаляются слоты, входящие в класс  $X$ , так как  $C'$ , будучи подклассом  $X$ , их унаследует.

**ИК19.** Класс  $C'$  добавляется в список подклассов класса  $X$ .

**ИК20.** Из класса  $C$  удаляются слоты, входящие в класс  $X$ , так как  $C$ , будучи подклассом  $X$ , их унаследует.

**ИК21.** Класс  $C$  добавляется в список подклассов класса  $X$ .

**ИК22.** Аналогично блоку **ИК12**, устанавливается, как взаимно расположены классы  $X$  и  $U$ . Для этого проверяется флаг  $F$ .

**ИК23.** Если флаг  $F$  сброшен, система не может принять однозначного решения относительно добавления класса  $X$  в иерархию класса  $U$ . Данную неоднозначность предлагается решить пользователю, для чего формируется вопрос: «Является ли  $X$  подклассом  $U$ ?». Если ответ пользователя отрицательный, алгоритм переходит к завершающей части цикла (блок **ИК27**).

**ИК24.** Если флаг  $F$  установлен, либо пользователь указал, что  $X$  является подклассом  $U$ , класс  $X$  добавляется в список подклассов класса  $U$ . Далее алгоритм переходит к завершающей части цикла (блок **ИК27**).



**ИК25.** Если классы  $C$  и  $C'$  различны, следует определить взаимное расположение классов  $C'$  и  $U$  проверкой флага  $F$ . Заметим, что в данном случае, если флаг  $F$  сброшен, нет оснований для предположений о взаимной связи классов  $C'$  и  $U$ , а различие  $C$  и  $C'$  дает возможность предположить обратное. Поэтому, если флаг  $F$  сброшен, алгоритм переходит к завершающей части цикла (блок ИК27).

**ИК26.** Если флаг  $F$  установлен, класс  $C'$  добавляется к списку подклассов  $U$ .

**ИК27.** Завершающая часть цикла – возможно, установленный флаг  $F$  сбрасывается.

**ИК28.** Итерирование по списку подклассов класса  $U$  – получение следующего элемента списка подклассов класса  $U$  и запись его в переменную  $C$ .

**ИК29.** Переход на начало цикла (блок ИК3) «Интеграция с подклассами  $U$ » с текущим значением итератора  $C$ .

**Пример работы алгоритма.** Пусть дано две онтологии, описывающие основные современные программно-аппаратные платформы персональных компьютеров. Базовая онтология (рис. 4) содержит следующие онтологические знания:

- существует две основных архитектуры процессоров: архитектура «RISC» и архитектура «x86»;
- существует два класса процессоров, отличающихся своей архитектурой: процессоры «PowerPC» с архитектурой «RISC» и процессоры «x86-совместимые» с архитектурой «x86»;
- существует класс компаний, производящих операционные системы: компании корпорации Microsoft (класс «Microsoft Corp.») и компании корпорации Apple (класс «Apple Inc.»);
- существует класс операционных систем, подразделяющийся на системы семейства «Windows» и «Mac OS», причем известно, что Windows-системы поддерживают базовые процессоры класса «x86-Совместимые» и производятся компаниями корпорации Microsoft, а системы Mac OS ориентированы на процессоры RISC и разрабатываются компаниями корпорации Apple.

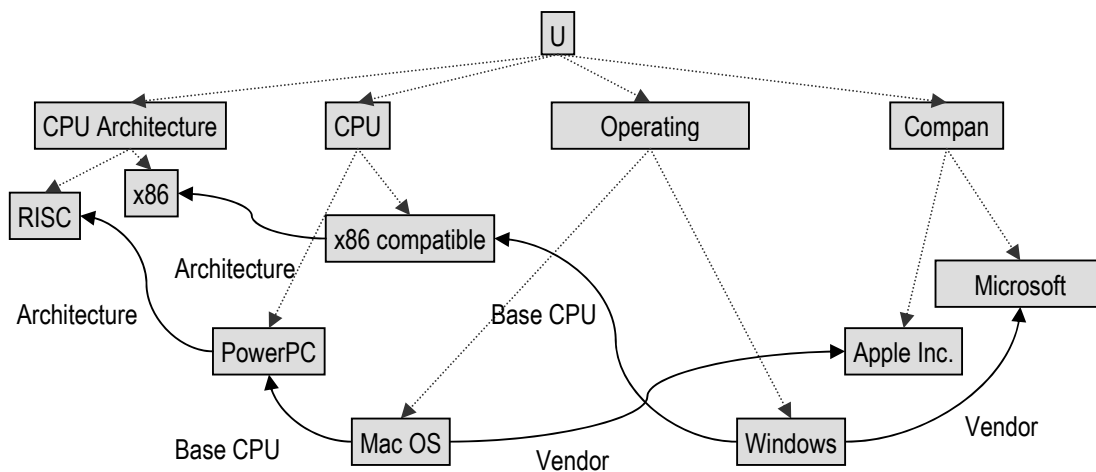


Рис. 4. Пример базовой онтологии

При поиске последней информации в данной ПдО поисковая система обнаружила и передала в модуль системной интеграции следующие онтологические знания в качестве интегрируемой онтологии (рис. 5):

- существует класс процессоров с архитектурой «x86», называемых семейством процессоров «Intel»;
- существует класс так называемых «Intel-based» операционных систем, поддерживающих базовые процессоры класса «Intel»;

– корпорацией Apple была разработана новая операционная система «Mac OS X» (точнее, класс операционных систем), принадлежащая классу «Intel-based» операционных систем.

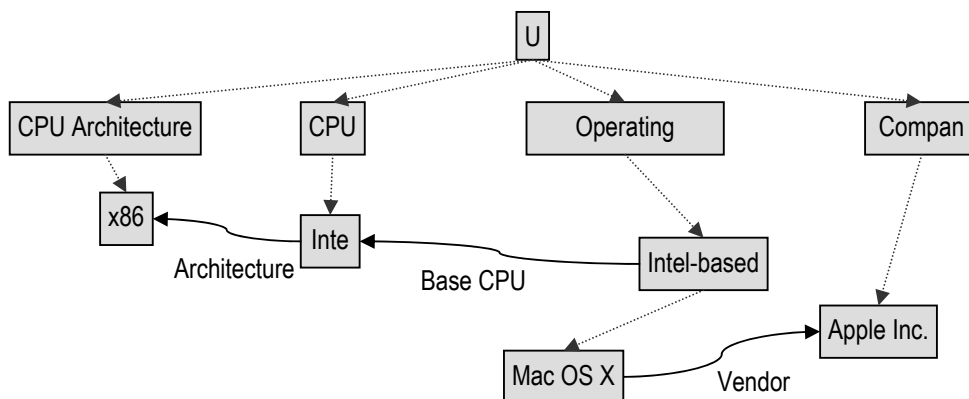


Рис. 5 Пример интегрируемой онтологии

В результате выполнения процедуры «Интеграция онтологий» базовая онтология будет дополнена онтологическими знаниями интегрируемой онтологии, изменив при этом и собственную исходную структуру, как это показано на рис. 6.

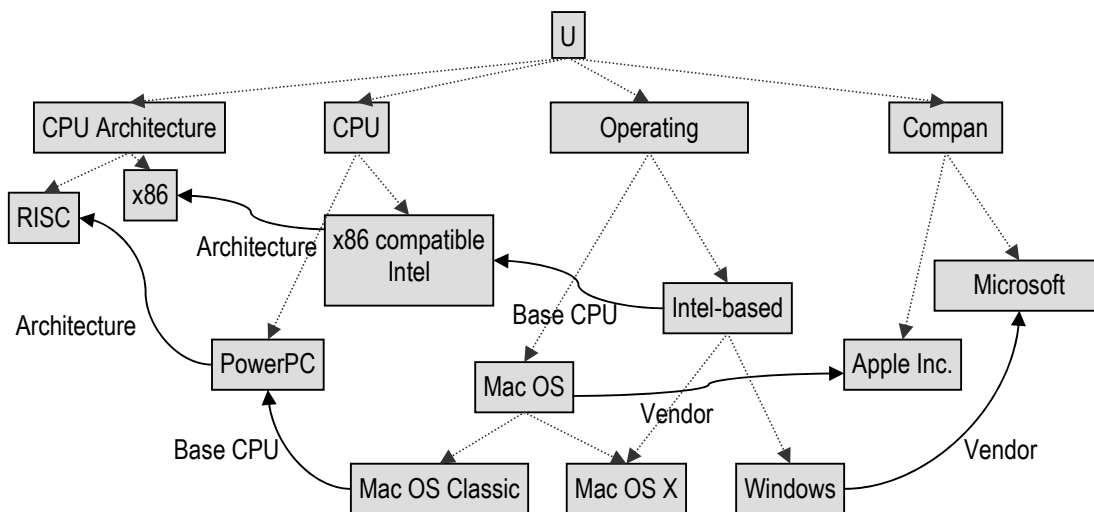


Рис. 6. Пример результирующей онтологии

**Сравнение результатов работы алгоритма с результатами работы PROMPT.** Произведя объединение двух данных онтологий средствами PROMPT, можно убедиться в значительных недостатках алгоритма его работы (см. рис. 7). Для объективной оценки алгоритма работы рассматриваемого плагина действия пользователя сводились к применению всех операций, предложенных PROMPT в том порядке, в котором они поступали, эмулируя тем самым автоматический режим работы.

Первым недостатком PROMPT является порядок определения пар классов для объединения или копирования – список операций в PROMPT строится в порядке добавления классов-аргументов операции в исходные онтологии. Очевидно, данный порядок является порядком хранения классов в онтологиях, то есть PROMPT не пытается предложить какую-либо «разумную» последовательность операций – он

выстраивает их «as is». Это, в свою очередь, привело к тому, что слоты в результирующую онтологию были добавлены до того, как были добавлены классы, на которые эти слоты ссылаются, что нарушило семантические связи между классами (на рис. 7 полностью отсутствуют семантические отношения).

Вторым недостатком PROMPT следует отметить способ сравнения классов. Несмотря на совпадение по содержанию классов «x86-compatible» и «Intel», они были скопированы в результирующую онтологию, так как имеют различные имена. Более того, в результирующей онтологии классы «Mac OS» и «Mac OS X» объединились по причине лексического подобия имен, что, несомненно, не соответствует действительности, так как данные операционные системы представляют абсолютно разные классы, отличающиеся типом базового процессора, что явно указано в исходных онтологиях.

Так же, не была установлена родовидовая связь между классами «Mac OS», «Mac OS X» и «Windows», так как PROMPT проводит сравнение классов с двумя вариантами результата сравнения («классы эквивалентны» или «классы различны») и не представляет никаких операций над классами, кроме объединения и копирования.

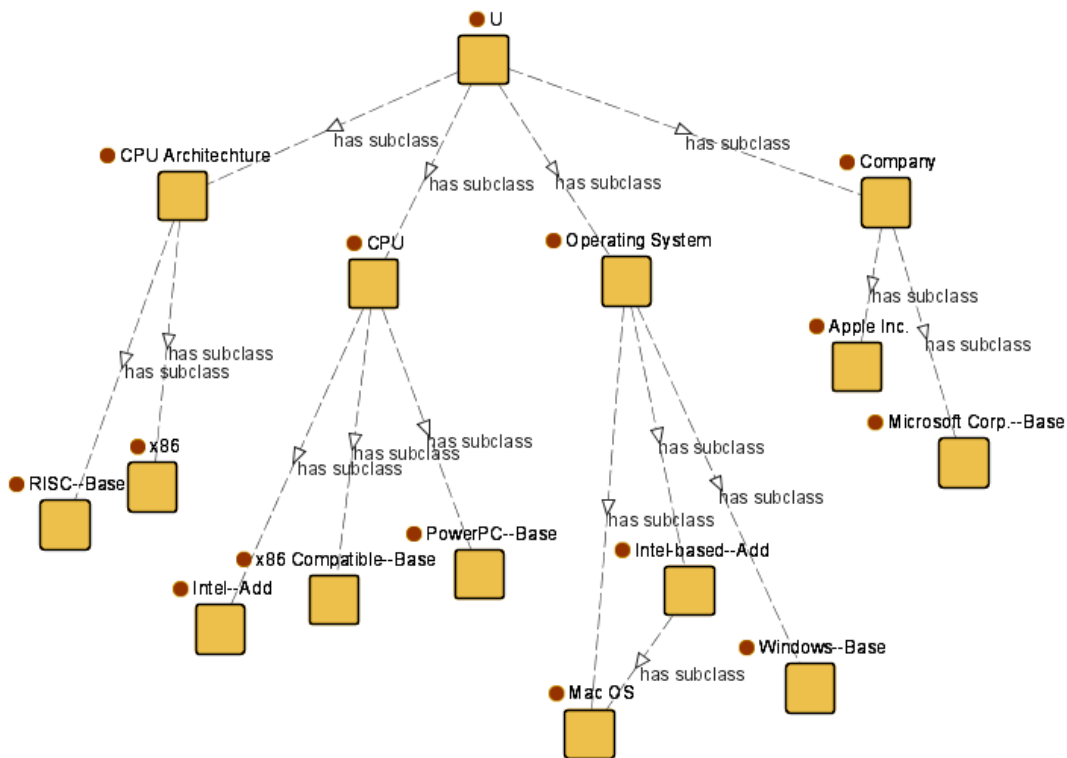


Рис. 7. Пример результирующей онтологии, полученной при помощи Protégé PROMPT

## Выводы

Рассмотрение содержания класса как его определяющей составляющей позволяет обобщить операцию сравнения классов от определения классической эквивалентности (варианты «классы эквивалентны» и «классы различны») до установления родовидовых отношений, таких как «классы  $C_1$  и  $C_2$  эквивалентны», «класс  $C_1$  является обобщением класса  $C_2$ », «класс  $C_1$  является уточнением класса  $C_2$ », «классы  $C_1$  и  $C_2$  частично эквивалентны» и «классы  $C_1$  и  $C_2$  различны»). Обобщенная операция сравнения классов, в свою очередь, расширяет набор возможных при объединении онтологий действий от двух («объединить» для эквивалентных классов и «скопировать» – для различных) до пяти («объединить», «создать пару класс-подкласс», «создать пару класс-надкласс», «создать общий надкласс

для данных двух классов» и «скопировать классы» в соответствии с результатами сравнения). Такой расширенный набор операций над сравниваемыми классами, вместе с оригинальным алгоритмом перебора, сравнения и выполнения операций над классами, качественно усовершенствуют процедуру объединения онтологий, реализованную в современных онторедаторах, превращая ее в процедуру системной интеграции онтологий предметных областей.

Дальнейшие исследования в области интеграции онтологий ПдО следует посвятить усовершенствованию операции сравнения классов путем введения в сравнение сложных функций интерпретации (элементов множества  $F$  модели  $\langle X, R, F \rangle$ ), использованию дополнительной информации об объемах классов, привлечению статистической информации о подобию классов, составленной на основе сравнения классов по содержанию. Целью исследований в этом направлении является приближение операции сравнения классов к уровню обработки глубокой семантики понятий, формирующих онтологии, тогда как полное достижение этого уровня является не разрешенной на сегодняшний день научной проблемой, что, очевидно, выходит за рамки данного исследования. Относительно процедуры интеграции онтологий, усовершенствования могут быть внесены в части преодоления возможных противоречий, вызванных введением в сравнение классов сложных функций интерпретации, визуализации процесса интеграции, предоставив пользователю возможность интерактивно следить за процессом, ускорения работы процедуры интеграции за счет оптимизации ее алгоритма. Сама разработанная процедура интеграции онтологий и алгоритм ее работы, описанные в данной статье, требуют дальнейшей экспериментальной проверки.

---

## Литература

---

- [Chaudhri 1998] V. Chaudhri, A. Farquhar, R. Fikes P. Karp J. Rice. OKBC: A Programmatic Foundation for Knowledge Base Interoperability. // Fifteenth National Conf. on Artificial Intelligence. AAAIPres/The MIT Press, Madison, P.600-607, 1998.
- [Dean 2004] Mike Dean and Guus Schreiber. OWL Web Ontology Language Reference. W3C Recommendation. – 2004.
- [Download Protégé] <http://protege.stanford.edu/download/download.html>
- [Ehrig 2005] Marc Ehrig, York Sure. FOAM – Framework for Ontology Alignment and Mapping; Results of the Ontology Alignment Initiative // Proceedings of the Workshop on Integrating Ontologies, CEUR-WS.org, 156. – 2005. pp. 72-76.
- [ISO 5807:1985] ISO 5807:1985. Information processing – Documentation symbols and conventions for data, program and system flowcharts, program network charts and system resources charts. International Organization for Standardization, 1985. – 25 p.
- [Knublauch 2004] Holger Knublauch, Ray W. Ferguson, Natalya F. Noy, Mark A. Musen. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications // Third International Semantic Web Conference – ISWC 2004, Hiroshima, Japan. – 2004. pp. 229-243.
- [McGuinness 2000] McGuinness D., Fikes R., Rice J., Wilder S. An environment for merging and testing large ontologies // In Proc. of the Seventh Int. Conf., KR2000, Morgan Kaufmann Publishers, San Francisco, CA. – 2000. pp. 483-493.
- [Noy 2000 (1)] Natalya Fridman Noy, Ray W. Ferguson, Mark A. Musen. The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility // Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management, Springer-Verlag London, UK. – 2000. pp. 17-32.
- [Noy 2000 (2)] Noy, N.F. and Musen, M.A. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. // Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000), Austin, TX. – 2000. pp 450-455.
- [Noy 2001] N. F. Noy, M. A. Musen. Anchor-PROMPT: Using Non-Local Context for Semantic Matching. // Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001), Seattle, WA. – 2001.
- [Noy 2003] N.F. Noy and M.A. Musen The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping // International Journal of Human-Computer Studies, 59/6. – 2003. pp. 983-1024.
- [Protégé project] Stanford Centre for Biomedical Informatics Research: Protégé project publications. <http://bmir.stanford.edu/publications/project.php/protg>

- [Sachs 2006] Eliza Sachs. Getting Started with Protégé-Frames. – 2006 – 72 p.
- [Войшвилло 2007] Войшвилло, Е.К. Понятие как форма мышления: логико-гносеологический анализ. М.: Изд-во ЛКИ, 2007. – 240 с.
- [ГОСТ 19.701-90] ГОСТ 19.701-90. Схемы алгоритмов, программ, данных и систем. Москва: Государственный комитет СССР по управлению качеством продукции и стандартам, 1990.
- [Ивлев 2001] Ивлев Ю.В. Логика. Учебник для вузов. 2-е изд., перераб. и доп. М.: Логос, 2001. – 272 с.
- [Овдей 2004] Овдей О.М., Проскудина Г.Ю. Обзор инструментов инженерии онтологий // Электронные библиотеки – Москва: Институт развития информационного общества, т.7 вып.4, 2004. – Электронный журнал, посвященный созданию и использованию электронных библиотек. <http://www.elbib.ru/>
- [Палагін 2006] Палагін О.В., Петренко М.Г. Модель категоріального рівня мовно-онтологічної картини світу // Математичні машини і системи. – 2006. – №3. – С. 91-104.
- [Палагін 2009] Палагін О., Кургаєв О. Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки // Вісник НАН України. – 2009. – № 3. – С. 14–25.

### 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

*Основные области научных исследований: системная интеграция трансдисциплинарных научных знаний, онтологический инжиниринг*



**Андрей Васильевич Михайлюк** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: fruler@ukr.net

*Основные области научных исследований: формальные модели представления знаний, логико-онтологические системы обработки знаний*



**Величко Виталий Юрьевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: velychko@aduis.com.ua

*Основные области научных исследований: индуктивный логический вывод, обработка естественно-языковых текстов.*



**Петренко Николай Григорьевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: petrng@ukr.net

*Основные области научных исследований: методология и инструментальные средства автоматизированного проектирования онтологий предметных областей, системная интеграция трансдисциплинарных научных знаний*

## КОМПЬЮТЕРНЫЕ ОНТОЛОГИИ И ОНТОЛОГО-УПРАВЛЯЕМАЯ АРХИТЕКТУРА ИНФОРМАЦИОННЫХ СИСТЕМ

Николай Петренко

**Abstract:** В работе описана архитектура двух основных компонент онтологической системы с обработкой естественно-языковых объектов – компьютерной онтологии и программно-аппаратных средств с онтолого-управляемой архитектурой.

**Keywords:** компьютерные онтологии, онтолого-управляемая архитектура.

**ACM Classification Keywords:** I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods.

---

### Введение

Системный подход к познанию ориентирует аналитика на рассмотрение любой ПДО с позиций закономерностей системного целого и взаимодействия составляющих его частей. Системность знаний исходит из многоуровневой иерархической организации любой сущности, то есть все объекты, процессы и явления можно рассматривать как множество более мелких подмножеств (признаков, деталей) и, наоборот, любые объекты можно (и нужно) рассматривать как элементы более высоких классов обобщений [Палагин, 2009, а].

Согласно [Аверкин, 1992] под *анализом* понимается вид исследования, при котором реальный или мыслимый объект расчленяется на составляющие части (элементы) и исследуются эти элементы и связи между ними. Анализ предметной области представляет особый вид научной деятельности, в результате которой строится интерпретационная модель предметных знаний (в широком смысле). В процессе анализа последние делятся на инвариантные и прагматические знания, концептуальные составляющие которых представляют онтологические знания ПДО.

---

### Компьютерные онтологии

Начало 90-х годов прошлого столетия считается зарождением *парадигмы компьютерных онтологий*. Она была сформулирована как попытка сгладить (и по возможности устранить) всё чаще проявляющиеся разного рода противоречия при функционировании и внедрении интеллектуальных систем с использованием баз знаний предметных областей. Ярким представителем таких систем на то время были экспертные системы (ЭС). Они успешно и эффективно функционировали в пределах одного коллектива и на уровне коммерческих образцов. У разработчиков постоянно возникал вопрос: “Как обеспечить продвижение этих образцов к конечному пользователю?”. Были предложены разработки оболочек – “пустых” ЭС и ряд других новшеств. Но они не имели решающего значения. Напрашивался вывод, что для эффективного функционирования ЭС на самом важном этапе “жизненного цикла” – функционирования у конечного пользователя при решении реальных задач – необходимо к каждой ЭС “приложить” эксперта в соответствующей ПДО. Необходимость присутствия эксперта объяснялась, в частности, быстротечностью изменения знаний во многих предметных областях и соответственно необходимостью обновления базы знаний ЭС в “реальном времени”. В тех же ПДО, где знания обладали относительной инвариантностью, ЭС продолжали эффективно функционировать.

Следующим этапом развития парадигмы компьютерных онтологий являлось задекларированное в [Guarino, 1998] построение ИС с онтолого-управляемой архитектурой (или онтолого-управляемых ИС).

Сказанное выше и ряд других факторов “подтолкнули” к разработке парадигмы компьютерных онтологий, основные принципы которой были сформулированы в [Gruber, 1993].

1. Доходчивость, ясность (Clarity). Термины (и понятия) онтологии должны отражать реальную действительность. Их символичные обозначения (знаки) должны формироваться на основе

- общепринятых правил в семиотике и должны выражать общепринятые смыслы реальных объектов. В свою очередь, эти смыслы извлекаются из общепринятых определений терминов (понятий), зафиксированных в толковых словарях, различных глоссариях ПдО. Суждения, входящие в определения, формализуются на основе формального общепринятого аппарата в виде тождественно истинных логических аксиом.
2. Обоснованность, связность (Coherency). Формирование начального набора понятий онтологии и их добавление должно быть обоснованным, определяемым, в первую очередь, требованиями предполагаемой совокупности решаемых задач. Логические аксиомы начального набора понятий должны быть непротиворечивыми. Для этого должен быть предусмотрен механизм логического вывода, который, в том числе, проверяет на непротиворечивость добавляемые аксиомы и выводимые в онтологии утверждения.
  3. Расширяемость (Extendibility). Ядром онтологии являются первоначально введенные (спроектированные) понятия и описывающие их аксиомы. В онтологии должен быть предусмотрен механизм расширения (ограничения) совместно используемых словарей понятий без нарушения целостности системы.
  4. Минимальное влияние кодирования (Minimal encoding bias). В онтологической системе (ОнС) должен быть реализован принцип совместного использования онтологий, который предполагает: спецификацию онтологии на уровне представления, а не символического кодирования; запись такой спецификации на общепринятом и платформонезависимом языке описания онтологий можно передать для использования любому программному агенту.
  5. Минимальные онтологические обязательства (Minimal ontological commitment). Этот принцип перекликается с принципами обоснованности и расширяемости/ограничения. Важно, чтобы множество понятий онтологии отображало концептуальную структуру ПдО, относительно стабильную на протяжении “жизненного цикла” ОнС. А последняя предоставляла возможность расширения или специализации отдельных ветвей онтологического графа (онтографа). Отделение концептуальных знаний от знаний, выраженных фактами, является стратегией построения ОнС, а точнее – онтологических баз знаний.

В отличие от онтологии как философской дисциплины остановимся на рассмотрении компьютерной (формальной) онтологии предметной области и онтологии домена предметных областей. Последняя (в том числе) имеет важное значение для объединения (интеграции) концептуальных знаний близких предметных областей или реализации технологии системной интеграции трансдисциплинарных научных знаний. Мы также выделяем *начальную онтологию ПдО*, являющуюся инициализирующей доминантой при реализации технологии автоматизированного построения онтологии ПдО.

Известны общепринятые определения онтологии ПдО Грубера, Гуарино, Гавриловой, на основе которых разные исследователи формулируют свои, частные определения онтологии ПдО – в соответствии с их конкретной областью профессиональных интересов<sup>1</sup>. Тематика и практическая направленность наших исследований (онтология как средство построения баз знаний и трансдисциплинарных научных исследований) предопределяет следующее определение компьютерной онтологии ПдО.

**Компьютерная онтология ПдО** – это:

- иерархическая структура конечного множества понятий, описывающих заданную предметную область;
- структура представляет собой онтограф, вершинами которого являются понятия, а рёбрами – семантические отношения между ними;
- понятия и отношения интерпретируются в соответствии с общезначимыми функциями интерпретации, взятыми из электронных источников знаний из заданной ПдО;

---

<sup>1</sup> Таких определений известно несколько десятков.

- дополнительные интерпретации понятий и отношений определяются аксиомами и ограничениями их области действия;
- формально онтограф описывается на одном из языков описания онтологий;
- функции интерпретации и аксиомы описаны в некоторой подходящей формальной теории.

Общепринято онтологию некоторой ПДО формально представлять упорядоченной тройкой [Палагин, 2005; Гаврилова, 2001; Guarino, 1998]

$$O = \langle X, R, F \rangle, \quad (1)$$

где  $X, R, F$  – конечные множества соответственно:  $X$  – концептов (понятий, терминов) предметной области,  $R$  – отношений между ними,  $F$  – функций интерпретации  $X$  и/или  $R$ .

Рассмотрение граничных случаев множеств (1):  $R = \emptyset$ ;  $R \neq \emptyset$ ;  $F = \emptyset$ ;  $F \neq \emptyset$  во всех четырёх комбинациях значений  $R$  и  $F$  даёт различные варианты онтологических конструкций, начиная от простого словаря и таксономии до формальной структуры концептуальной базы знаний для высокоинтеллектуальных знание-ориентированных систем. По своей функциональной полноте и степени формальности различаются три вида онтологий: простая, полная (или строгая) и множество промежуточных или неполных онтологий.

*Простая онтология* – это такая онтология, в которой  $R = \emptyset$ ;  $F = \emptyset$ . Она служит (в основном) для однозначного восприятия научным сообществом понятий в соответствующей прикладной области.

*Строгая* или *полная онтология* ( $R \neq \emptyset$ ;  $F \neq \emptyset$ ) – это такая онтология, в которой множества концептов и концептуальных отношений максимально полные, а к функциям интерпретации добавляются аксиомы, определения и ограничения. При этом описания всех компонент представлены на некотором формальном языке, доступном для их интерпретации компьютером. Схема формальной модели полной онтологии описывается четвёркой:

$$O = \langle X, R, F, A(D, R_s) \rangle, \quad (2)$$

где  $X$  – множество концептов;

$R$  – множество концептуальных отношений между ними;

$F : X \times R$  – конечное множество функций интерпретации, заданных на концептах и/или отношениях;

$A$  – конечное множество аксиом, которые используются для записи всегда истинных высказываний (определений и ограничений);

$D$  – множество дополнительных определений понятий;

$R_s$  – множество ограничений, определяющих область действия понятийных структур.

Полная (или компьютерная) онтология является (формальным) выражением концептуальных знаний о предметной области и по своей значимости сопоставима с базой знаний знание-ориентированной информационной системы (ЗОИС), а её построение является специфической формой человеческого творчества. Творческий процесс можно представить совокупностью операций-процедур с суждениями, утверждениями, понятиями и отношениями между ними. А последние являются фундаментом, основой для построения составной части научной теории – онтологической базы знаний в заданной предметной области. При этом такие знания описываются в декларативной форме.

Известно, что проектируемые средства информатики в соответствии с их проблемной ориентацией базируются на некоторой совокупности фундаментальных принципов, методик и алгоритмов. Также известны многочисленные разработки в области обработки ЕЯТ [Широков, 2005 и др.], в которой ввиду огромной сложности решаемых задач принимаются те или иные ограничения, снижающие эффективность средств обработки. Автором были рассмотрены и проанализированы разработки, в которых формальная



онтология играет существенную роль. В результате были выявлены главные недостатки, снижающие эффективность реализации отдельных процедур и конечных результатов в целом.

Известны общие совокупности методов, принципов, процедур и атрибутов системного анализа как научного познания в любой ПдО [Згуровский, 2005 и др.]. При конкретизации ПдО конкретизируются и средства системного анализа. Парадигма компьютерных онтологий, развиваемая во взаимодействии со средствами и методами системного анализа, положила начало развитию новой ветви средств и методов системного анализа ПдО – системно-онтологического анализа (или подхода).

Центральной идеей системно-онтологического подхода (СОП) является разработка онтологических средств поддержки решения прикладных задач – полифункциональной онтологической системы. Такая система (точнее, её концептуальная часть) описывается двойкой (3) и представляет онтологию ПдО, состоящую из онтологии объектов и онтологии процессов, и онтологию задач [Палагин, 2009, а].

$$O_{HC} = \langle O^{поо}(O^O, O^П), O^З \rangle \quad (3)$$

На рис. 1 представлена схема онтологий-компонентов предметной области и проблемного пространства (ПрП). ПрП – это модель всех таких аспектов или компонент ПдО, с которыми связаны (опосредственно или непосредственно) знания, требующиеся при решении различных задач в этой ПдО. Всякое ПрП состоит из двух блоков: инвариантной (относительно неизменной) части и множества изменяемых частей, соответствующих отдельным задачам. В составе инвариантной части, например в методологии SMEE (Structured Methodology for Elicitation of Expertise), выделяют семь типов компонент: объекты, инструменты, операторы, операции, конечные продукты, побочные продукты и ограничения [Методический комплекс]. Эти типы компонент – суть понятия, которые хорошо группируются в онтологии объектов и процессов, представленных на рис. 1. На нём приняты следующие обозначения:

- $O^O$  – онтология множества объектов (понятий, концептов) ПдО, которая рассматривается как иерархическая структура классов, подклассов и элементов классов;
- $O^П$  – онтология множества процессов ПдО, которая рассматривается как иерархическая структура процессов, подпроцессов, действий и операций;
- $O^З$  – онтология совокупности задач (типовых наборов), которые могут быть поставлены и решены в ПдО. Рассматривается как иерархическая структура задач, подзадач, процедур и операторов.

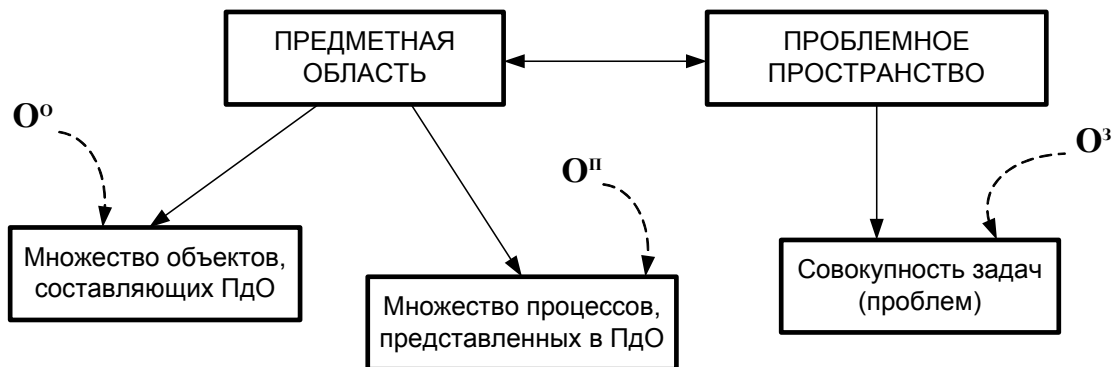


Рис. 1. Схема онтологий-компонентов предметной области

Функционально полная (с практической точки зрения) компьютерная онтология ПдО содержит декларативные ( $O^O$  и  $O^П$ ) и процедурные ( $O^З$ ) знания. Эти знания, взаимодействуя между собой, обеспечивают знание-ориентированную поддержку процесса решения типового набора задач пользователя. Инициальными являются процедурные знания, а точнее – компоненты онтологии задач – обобщенный алгоритм решения типового набора задач (представленный в онтологическом виде), формирователь методов решения типового набора задач и блок решателя задач. После того, как выбран метод и адаптирован алгоритм для решения целевой задачи, блок решателя задач формирует исходные

данные и последовательно запускает соответствующие процессы из онтологии процессов, которая, в свою очередь, обращается к соответствующим концептам онтологии объектов.

Описанное выше взаимодействие онтологии ПдО и онтологии ПрП является одной из основных характеристик онтолого-управляемости ЗОИС, другие же характеристики приведены в [Палагин, 2009, б].

### Онтолого-управляемая архитектура ИС

В многочисленных зарубежных публикациях под *онтолого-управляемой архитектурой* (Ontology-driven architecture) понимают такую архитектуру ЗОИС, в которой база знаний ПдО представлена формальной онтологией. При этом последняя описана на некотором языке описания онтологий, а список таких языков определён консорциумом W3C (OWL, RDFS и др.). ЗОИС с онтолого-управляемой архитектурой классифицируют [Палагин, 2005] по функциональному признаку на онтолого-знающие информационные системы (ОЗнИС) и онтолого-управляемые информационные системы (ОУИС). ОЗнИС характеризуется “пассивной” ролью онтологии, т.е. выполнением ею роли только информационно-справочного компонента. Любая другая подсистема только “знает” о существовании (возможно отдалённо) онтологии и может использовать её для любой специфической цели приложения. В ОУИС онтология играет “активную” роль, кооперируется с другими подсистемами в текущем времени в соответствии с “высшей” общей целью ЗОИС. Важным следствием использования онтологии в текущем времени является способность коммуникации между программными агентами. Последние поддерживают связь друг с другом через сообщения, которые содержат выражения, сформулированные в терминах онтологии (*онтолого-управляемая коммуникация*). Для того чтобы программному агенту “понять” значения этих выражений, ему необходим доступ к онтологии.

Известна и другая трактовка онтолого-управляемой архитектуры. Это новая концепция разработки программ – MDD (Model Driven Development – разработка, управляемая моделями) и вариант ее реализации – MDA (Model Driven Architecture – архитектура, управляемая моделью, или модельно-управляемая архитектура), предполагающая разработку архитектуры программного комплекса, ориентированную на предметную область и не зависящую от конкретной платформы программирования. Развитием данной концепции является подход к разработке программного обеспечения, предлагаемый консорциумом W3C для разработки, в частности, серверных приложений и называемый архитектурой, управляемой онтологией. [www.w3.org].

Выше приведено общепринятое (по крайней мере, в зарубежных публикациях) описание онтолого-управляемой архитектуры для ИС различного назначения. Для ЗОИС с обработкой ЕЯО оно является фундаментальным, основано на свойстве активности знаний, в том числе онтологических. Специфика обработки языковых и предметных знаний предполагает включение в архитектуру ОУИС соответствующих онтологий, которые в свою очередь упорядочиваются в соответствии с онтологической иерархией. И главной особенностью архитектуры ОУИС, подчёркивающей её онтолого-управляемость, является явное представление процедурных знаний в виде онтологии задач [Палагин, 2009, а].

На рис. 2 представлена схема фрагмента архитектуры ОУИС, отображающая только её “онтологическую” компоненту. Это “ядро” системы, между которым устанавливаются эксплицитные связи с процессами и процедурами (или технологической цепочкой) ИИТ обработки знаний, содержащихся в ЕЯО, с приложениями (через онтологию задач), с научной (онтология верхнего уровня) и языковой (языково-онтологическая) картинами мира. На нём приняты следующие обозначения:  $ОД_k$  ПдО – онтология  $k$ -го домена предметных областей, где  $k = \overline{1, K}$ ,  $K = Card OD$  – мощность множества онтологий доменов, интегрированных в библиотеку;  $О ПдО_i$  – онтология  $i$ -ой предметной области, где  $i = \overline{1, I}$ ,  $I = Card ПдО$  – мощность множества онтологий предметных областей, интегрированных в библиотеку;  $ОЗ_j^i$  – онтология  $j$ -го типового набора задач  $i$ -ой предметной области, где  $j = \overline{1, J}$ ,  $J = Card OZ$  – мощность множества типовых наборов задач  $i$ -ой предметной области;

ЛО ПдО<sub>*i*</sub> – лингвистическая онтология *i*-ой предметной области; ЯОКМ – языково-онтологическая картина мира; 1 – обработка омонимичных форм; 2 – построение поверхностных семантических структур; 3 – обработка семантических неоднозначностей; 4 – обработка концептуальных структур; ИИТ – интегрированная информационная технология обработки языковых и предметных знаний.

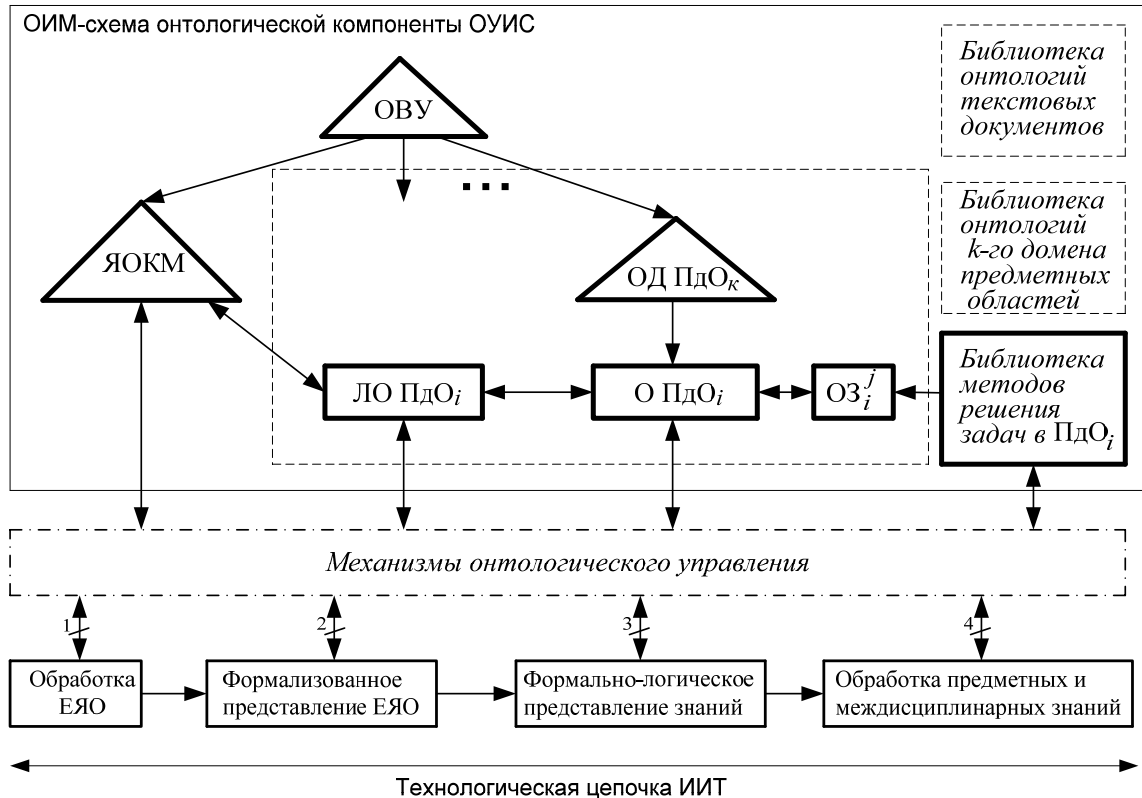


Рис. 2. Фрагмент архитектуры “онтологической” части ОУИС с обработкой ЕЯО

Итак, ЗОИС с онтолого-управляемой архитектурой присущи следующие характерные черты:

композиция онтологий разного уровня и назначения, как по вертикали, так и по горизонтали. По вертикали интегрируются онтология верхнего уровня, онтология домена предметных областей и онтология предметной области. По горизонтали интегрируются онтологии предметных знаний и знаний проблемного пространства. В свою очередь, предметные знания взаимодействуют с языковыми знаниями через лингвистическую онтологию предметной области;

- эффективное многократное использование онтологии предметной области и онтологии задач для разных наборов типовых задач. Будучи один раз построенными указанные онтологии пригодны для решения произвольного набора задач. При этом для новой задачи только составляется её расширенная спецификация на естественном языке, которая затем анализируется лингвистическими средствами ОУИС с обработкой ЕЯО. Результат анализа передаётся в блок решателя задач, где формируются обобщённый алгоритм решения задачи, наборы понятий-объектов и понятий-процессов, участвующих в решении задачи, и выбираются методы решения задачи;
- следствием композиции принципов и механизмов двойной парадигмы предметных знаний и онтологического управления является интеграция и эффективное применение компонентов архитектуры ОУИС в архитектурно-структурную организацию **интегрированных инструментальных средств автоматизированного построения онтологических баз знаний предметных областей**;

- частным случаем описанного в предыдущем пункте следствия является следующее. Произвольная онтология (простая, смешанная или формальная) эффективна, прежде всего, в обучении. Действительно, гораздо эффективнее показать новому сотруднику предприятия онтограф структурных подразделений предприятия, чем долго рассказывать о той же самой структуре. И только ИС с онтолого-управляемой архитектурой, имеющая в своём составе формальную онтологию некоторой предметной дисциплины, позволяет реализовать все преимущества электронного образования, в том числе и дистанционного обучения. Наиболее полно указанные свойства проявляются при использовании и взаимодействии онтологий двух уровней – онтологии домена предметных дисциплин (это может быть множество дисциплин, читаемых на кафедре или факультете) и онтологий самих предметных дисциплин. При этом естественным образом находят своё решение такие проблемы как: открытости и закрытости систем обучения; единообразия представления понятийных структур обучения; автоматизированного построения новых курсов обучения с учётом накопленных преподавателями кафедры знаний; замена кадрового состава преподавателей и др.

---

### Выводы

В работе рассмотрены основные принципы парадигмы компьютерных онтологий, описана схема онтологий-компонент предметной области и онтологическая часть архитектуры знание-ориентированной информационной системы с обработкой ЕЯО, обладающая признаками онтолого-управляемости. Дальнейшее направление исследований связано с решением одной из главных научно-технических задач разработки ЗОИС с обработкой ЕЯО – построение инструментальных средств автоматизированного проектирования онтологии ПДО на основе больших объемов текстовой информации.

---

### Литература

- [Аверкин, 1992]. Толковый словарь по искусственному интеллекту / Авт.-сост. А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1992. – 256 с.
- [Гаврилова, 2001]. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
- [Методический комплекс]. Методический комплекс по дисциплине “Современные проблемы науки” // Методология анализа предметных знаний. – Доступно на [http://old.ulstu.ru/people/SOSNIN/umk/Modern\\_Scientific\\_Problems/beloborodov/item\\_znan.htm](http://old.ulstu.ru/people/SOSNIN/umk/Modern_Scientific_Problems/beloborodov/item_znan.htm).
- [Палагин, 2005]. Палагин А.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: УНІВЕРСУМ. – 2005. – 680 с.
- [Палагин, 2009, а]. Палагин А.В., Петренко Н.Г. Системно-онтологический анализ предметной области // УСиМ. – 2009. – № 4. – С. 3–14.
- [Палагин, 2009, б]. Палагин А.В., Крытый С.Л., Петренко Н.Г. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация // УСиМ. – 2009. – № 3. – С. 42–55.
- [Широков, 2005]. Широков В.А., Бугайов О.В., Грязнухина Т.О. та ін. Корпусна лінгвістика. – К.: Довіра, 2005. – 471 с.
- [Guarino, 1998]. Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.) Formal Ontology and Information Systems. //Proceedings of FOIS'98. – Trento, Italy. – 1998. – 6–8 June. – IOS Press, Amsterdam. – PP. 3–15.
- [Gruber, 1993]. Gruber, T. R. 1993. A translation approach to portable ontology specifications // Knowledge Acquisition, 5 (2): PP. 199–220.
- [www.w3.org]. <http://www.w3.org/2001/sw/BestPractices/SE/ODA/>.

---

### Информация об авторе



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

Область исследований: методология и инструментальные средства автоматизированного проектирования онтологий предметных областей, системная интеграция трансдисциплинарных научных знаний

## ИСПОЛЬЗОВАНИЕ СПЕЦИАЛИЗИРОВАННОЙ ЛЕКСИЧЕСКОЙ ОНТОЛОГИИ ДЛЯ АВТОМАТИЗАЦИИ ФОРМИРОВАНИЯ ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ ПО ЕСТЕСТВЕННО-ЯЗЫКОВЫМ ТЕКСТАМ

Ольга Лесько, Юлия Рогушина

**Аннотация:** Рассматриваются методы автоматизации формирования онтологии предметной области (ПрО), интересующей пользователя, путем лингвистической обработки естественно-языковых документов, pertinentных этой ПрО, в частности, словарных статей толковых словарей, доступных через Web энциклопедий и т.д. Предлагается одновременно формировать специализированную лексическую онтологию (СЛО) этой ПрО, имеющую фиксированную структуру и содержащую сведения о лексемах, соответствующих терминам этой ПрО. СЛО затем используется для лингвистического анализа ЕЯ-текстов, позволяя обнаруживать фрагменты ЕЯ-текста, связанные с понятиями ПрО. Результаты такого анализа используются далее как для дополнения онтологии ПрО, так и для решения задач пользователя.

**Keywords:** Semantic Web Technologies, ontology, natural language processing

**ACM Classification Keywords:** I.2.4 Knowledge Representation Formalisms and Methods - Semantic networks, H.3.3 Information Search and Retrieval - Retrieval models

---

### Введение

Развитие информационных технологий и увеличение объемов обрабатываемых сведений привело к появлению более сложных методов представления и преобразования информации. В настоящее время большую распространенность получили информационные системы (ИС), базирующихся на знаниях, т.е. системы, в которых поддерживается обработка информации на семантическом уровне.

В частности, во многих ИС для представления распределенных, совместно используемых знаний используются онтологии и тезаурусы [1,2]. Построение онтологий различных областей, содержащих основные понятия и зависимости между ними, невозможно автоматизировать полностью, однако можно значительно снизить трудоемкость построения понятийного аппарата предметной области (ПрО) при помощи лингвистической обработки естественно-языковых текстов, релевантных этой ПрО и содержащих знания о ней. Подобные подходы широко используются при создании прикладных поисковых систем [3,4], систем дистанционного обучения [5] и электронной коммерции. В настоящее время уже разработаны системы автоматизированного построения онтологий, например Text2Onto [6].

Способы создания и использования онтологий в ИС очень сильно различаются в зависимости от направленности самих ИС и специфики ПрО, от того, какие именно естественные языки являются средством представления знаний о ПрО, но в основном они тем или иным образом ориентированы на стандарты и программные средства, разработанные в рамках проекта Semantic Web.

При разработке онтологии ПрО (ОПрО), интересующей пользователя, в данной работе предлагается использовать *специализированную лексическую онтологию* (СЛО) – такую лексическую онтологию этой ПрО, которая создается и пополняется параллельно разработке ОпрО на основе анализа корпуса ЕЯ-текстов, которые пользователь посчитал относящимися к исследуемой ПрО. СЛО является инструментом, на основе которого итеративно выполняется анализ семантики ЕЯ-текстов.

Принципиальное отличие СЛО от обычных лексических онтологий заключается в том, что в качестве экземпляров класса «лексема» в нее включаются только те понятия, которые связаны с понятиями ОпрО, а не все встречающиеся в ЕЯ понятия. Это значительно уменьшает размер онтологии и, соответственно, сокращает время ее обработки.

При решении ряда интеллектуальных задач пользователю надо использовать формализовано представленные знания соответствующей предметной области (ПрО), представленные в виде онтологии. Обычно такая онтология, полностью соответствующая информационным потребностям пользователя, им самим и должна быть построена с использованием фрагментов других онтологий и средств автоматизированного извлечения онтологических знаний из ЕЯ-текстов. Вспомогательным средством при этом является создание соответствующей лексической онтологии, которая в дальнейшем может использоваться и для решения других задач.

---

### Постановка задачи

---

Чтобы автоматизировать построение онтологии предметной области, предлагается одновременно с ее построением строить специализированную лексическую онтологию этой ПрО и использовать ее затем для извлечения знаний из ЕЯ-текстов: СЛО позволяет семантически размечать ЕЯ-текст понятиями ПрО и обнаруживать связи между этими понятиями, которые затем фиксируются в онтологии ПрО. Реализация базовых процедур анализа, синтеза и понимания естественного языка компьютером является первым необходимым звеном в цепочке «Компьютерная обработка естественно-языковых текстов – Представление знаний – Компьютерная обработка знаний» [3].

---

### Построение онтологии ПрО

---

Под онтологией (О) будем понимать упорядоченную тройку  $O = \langle X, R, F \rangle$ , где  $X$  - конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология  $O$ ;  $R$  - конечное множество отношений между концептами заданной предметной области;  $F$  - конечное множество функций интерпретации, заданных на концептах и отношениях онтологии  $O$  [7]. Основными типами онтологических отношений являются отношение принадлежности (**instance-of**) и отношение класс - подкласс (**subkind-of**). [8]. Отношения описывают связи между классами и экземплярами классов.

В контексте информационных технологий под онтологией понимается система или схема понятий (а не терминов) области знаний, отношений между ними и правил операций над ними, исчерпывающим образом представляющая область знаний. Онтология включает формальные определения всех используемых терминов (дескрипторов понятий), отношений и операций. Такие определения должны быть с одной стороны понятны разработчикам и, главное, пользователям, с другой стороны быть операбельными при машинной обработке. Следовательно, необходимо установить соответствие между словами естественного языка, понятными пользователю, и их формальными определениями, пригодными для автоматизированной обработки различными программами и алгоритмами.

Таким образом, возникает необходимость в построении двух онтологий:

- **СЛО**, под которой в дальнейшем мы будем понимать онтологию, понятиями которой является то подмножество лексем какого-либо ЕЯ, которое используется для описания понятий определенной ПрО, ее отношениями - онтологические отношения между лексемами (синонимия, меронимия, гипонимия, гиперонимия, антонимия, родство и т.д. [9]), экземплярами понятий могут быть поименованные сущности, а свойствами понятий - те словоформы ЕЯ, которые соответствуют лексеме, а множество аксиом является пустым)
- **онтологии предметной области**, в которой явным образом указываются связи между понятиями ПрО (при этом для определения понятий и связей используется терминология, формализованная в лексической онтологии).

Лексическая онтология может быть в дальнейшем использована при извлечении информации из ЕЯ-текстов. Такая онтология позволила бы извлекать информацию независимо от того, какими словами она сформулирована. Кроме того, она была бы полезна при решении таких задач, как разрешение эллипсисов, анафорических ссылок, морфологической и лексической омонимии.

В процессе построения онтологии ПрО предлагается выполнять лингвистический анализ текстов статей толковых словарей как данной ПрО, так и общей лексики. При этом словарные определения могут редактироваться пользователем в зависимости от его потребностей. Важно только, чтобы словарные статьи были хорошо отформатированными текстами, удовлетворяющими требованиям к определениям. Эти требования таковы:

Онтология ПрО позволяет формализовано представлять знания этой ПрО и использовать смысловые связи между понятиями, специфичные именно для этого домена (причем такие связи могут также описываться понятиями из лексической онтологии - теми словоформами, которые соответствуют глаголам и отглагольным существительным).

### Структура онтологии ПрО.

Множество концептов предметной области  $X$  состоит из двух подмножеств:  $X_1$  – множества концептов верхнего уровня, которые не определяются через другие концепты, и  $X_2$  – множества производных концептов, которые определяются через другие концепты:  $X = X_1 \cup X_2$ .

Концепты верхнего уровня  $X_1$  – это те знания, которые мы считаем априорно известными каждому пользователю, и, следовательно, нет необходимости формализовывать их каждый раз заново при создании каждой прикладной онтологии (например, онтологии ПрО). К таким знаниям относятся, например, сведения об единицах измерения («минута состоит из 60 секунд», географических объектах («Киев – столица Украины» и т.д. в качестве формализации таких знаний в дальнейшем мы предлагаем использовать ссылки на онтологии верхнего уровня или общие онтологии (такие, как SUMO[9]). Значительно сложнее сформировать множество  $X_2$ , уникальное для каждой ПрО. Каждый элемент множества  $X_2$  представляет собой вектор, состоящий из идентификатора класса ID, идентификатора суперкласса  $s$ , списка атрибутов  $A$ , и списка значений атрибутов  $V$ .

$$X_2 = \{x_i : x_i = \langle ID, s \in X, s \neq x_i, A = \{a_1, \dots, a_m\}, V \rangle\}$$

Множество отношений предметной области также представляет собой объединение множества атрибутов  $A$  и множества отношений  $R_0$ :  $R = A \cup R_0$ . Множество  $R_0$  в свою очередь состоит из множества отношений верхнего уровня, которые не определяются через другие концепты, и множества производных, которые определяются через другие концепты.

Множество понятий верхнего уровня можно представить, например, так

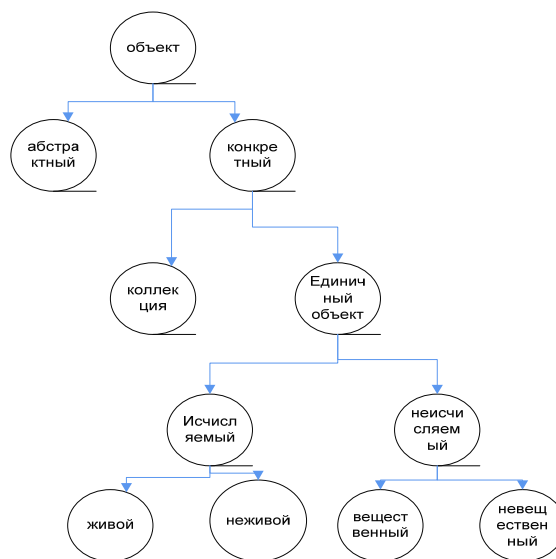


Рис.1. Пример онтологии верхнего уровня

Данная классификация построена на основе грамматических классов существительных и онтологии SUMO [10].

---

### Основные типы отношений между понятиями ПрО и методы их определения

---

Основные типы онтологических отношений и способы их выражения в естественном языке описаны, например, в [9] и [11].

Как указывается в [1] и [13], для определения различных видов отношений между словами в ЕЯ-текстах не всегда необходимо выполнять полный синтаксический анализ с учетом окончаний всех слов. В процессе речевого общения человек в первую очередь использует семантику слов, и только в случае неоднозначности используются окончания ([13], [15]) поэтому для выделения определительных отношений не предполагается использовать формальное описание синтаксических структур. В данном случае используются определения, данные пользователем, или найденные в «Великому тлумачному словнику Української мови» [17].

В некоторых случаях предлагается заменять существительные, выполняющие функцию определения, на соответствующие классы.

Например, в процессе анализа предложения «До цієї книги цікавість викликати важко» находим в словаре определение

ЦІКАВІСТЬ, -вості, ж. 1. Бажання, намагання дізнатися про щось в усіх подробицях; увага до кого-, чого-небудь; допитливість.

Так как в данном предложении есть только одно слово, которое может быть связано с существительным в виде «до кого-, чого-небудь», формируется словосочетание «цікавість до цієї книги».

---

### Структура специальной лексической онтологии (СЛО)

---

Структура лексической онтологии определяется ее назначением. Поскольку лексическая онтология должна отображать лексемы естественного языка в понятия ПрО, каждой лексеме должен быть приспан ID соответствующего класса, а также следующие виды валентности лексемы:

**лексическая валентность** - потенциальная сочетаемость с другими словами, лексическими группами или классами слов;

**морфологическая валентность** — потенциальная сочетаемость с морфемами словообразования и словоизменения. Под **синтаксической валентностью** — способность занимать определенные позиции в тех или иных синтаксических структурах и вступать в синтаксические связи с теми или иными классами слов.

Под лексемой [19] мы понимаем объединение корневых и аффиксальных морфем, образующих лексическую единицу, независимо от возможных для нее синтаксических функций, парадигмы и валентности.

Множество синтаксических правил отображает определенные синтаксические структуры в отношении ПрО с учетом семантики. Онтологические отношения (синонимия, таксономия) находятся в процессе обращения к онтологии ПрО.

---

### Общая схема построения СЛО в процессе разработки онтологии ПрО

---

Очевидно, что пользователю сразу и вручную создать онтологию ПрО, которая будет адекватно отражать все доступные пользователю знания об этой ПрО, – задача практически нереальная для любой сколько-нибудь сложной сферы деятельности. Поэтому следует строить такую онтологию итеративно, а в качестве вспомогательного средства формализации знаний использовать СЛО, в которую помещать информацию, обнаруженную пользователем в процессе анализа ЕЯ-текстов, связанных с этой ПрО. СЛО также



формируется пользователем не сразу, а итеративно, причем обе создаваемые и совершенствуемые онтологии оказывают влияние друг на друга.

Общая архитектура системы соответствует архитектуре системы автоматизированного построения онтологий Text2Onto [6]. Основными модулями системы являются on-line словарь, лингвистический процессор, и модуль пополнения онтологии.

При этом надо выполнить следующие шаги:

сформировать начальный словарь терминов ПрО. Терминами ПрО могут считаться лексемы, определения которых даются в текстах из данной ПрО, либо названия объектов, свойства которых описываются. Для нахождения определений в тексте используется алгоритм, приведенный в [18]. Например, терминами в ПрО «Мебель» могут быть «столик», «компьютерный столик», если даются их определения или свойства объектов. Определение слова «компьютерный» в данном случае не требуется, независимо от частоты его встречаемости. Такие определительные словосочетания находятся в результате семантико-синтаксического анализа, в процессе которого используется толковый словарь, как описано выше. Также термины могут вводиться вручную пользователем (при этом можно использовать подмножества и пересечения ранее сформированных наборов терминов, например, толковых словарей). Кроме того, для нахождения таких терминов можно использовать процедуру, приведенную в [13].

дать ЕЯ-определение каждого из выбранных пользователем терминов (например, "Кóшка — одомашненная форма мелкого хищного млекопитающего семейства кошачьих — дикой ближневосточной кошки" (самостоятельно либо воспользоваться различными толковыми словарями), сформировав таким образом корпус текста для дальнейшего анализа (при этом термины используются как ключевые слова при поиске в Web)

Определим более детально эту процедуру. Предположим, пользователь внес термин X в начальный словарь терминов ПрО. Если пользователь достаточно четко представляет себе его значение, то он пытается сам дать ему то определение, которое наиболее адекватно описываемой ПрО. Если пользователь не смог дать определение слову X, он находит некоторые определения X в толковых словарях, Википедии и т.д., выбирая устраивающие его. Если этого оказывается недостаточно, то пользователь обращается к поисковым системам Интернета с X в качестве ключевого слова и выбирает подходящие ИР, найденные этими ИПС. Затем все найденные результаты объединяются в один ЕЯ-документ – корпус текстов.

проанализировав выбранный заранее корпус текстов, расширить начальный словарь терминов ПрО теми понятиями, которые использовались в определениях, и опять найти ЕЯ - определения каждого из терминов (аналогично п.2 либо через ссылки на онтологии верхнего уровня)

эти шаги повторять итеративно до тех пор, пока пользователь не будет удовлетворен предоставленным ему набором понятий, а все термины будут определены либо через другие понятия, либо через ссылки на понятия онтологии верхнего уровня (ссылка типа «общеизвестный термин») или просто сопровождаемы указателем «неопределимо в рамках ПрО».

перейти к формированию СЛО. Элементы словаря терминов ПрО – это экземпляры класса «лексема» СЛО. Необходимо на основе анализа ранее сформированного корпуса текстов зафиксировать допустимые словоформы этих лексем, а также их морфологические свойства. На этом этапе для каждой из морф определяется класс слов, с которым она сочетается. Если таких классов больше одного, пользователю предлагается определить новый тип отношений. Каждой лексеме поставить в соответствие ID класса в онтологии ПрО.

перейти к формированию онтологии ПрО. Каждой лексеме из СЛО ставится в соответствие единственное понятие в онтологии ПрО. При этом нескольким лексемам из СЛО может соответствовать одно понятие онтологии ПрО

сформировать экземпляры класса «предложение» СЛО путем лексического анализа корпуса текстов.

установить семантические связи между теми понятиями онтологии ПрО, которым соответствуют лексемы, являющиеся свойствами одного экземпляра класса «предложение»,  
используя СЛО, осуществить семантическую разметку текста определений понятиями онтологии ПрО  
Проанализировать полученные онтологии, внести в них изменения и вернуться к п.7.  
Общая схема первого этапа алгоритма приведена на рис.2.

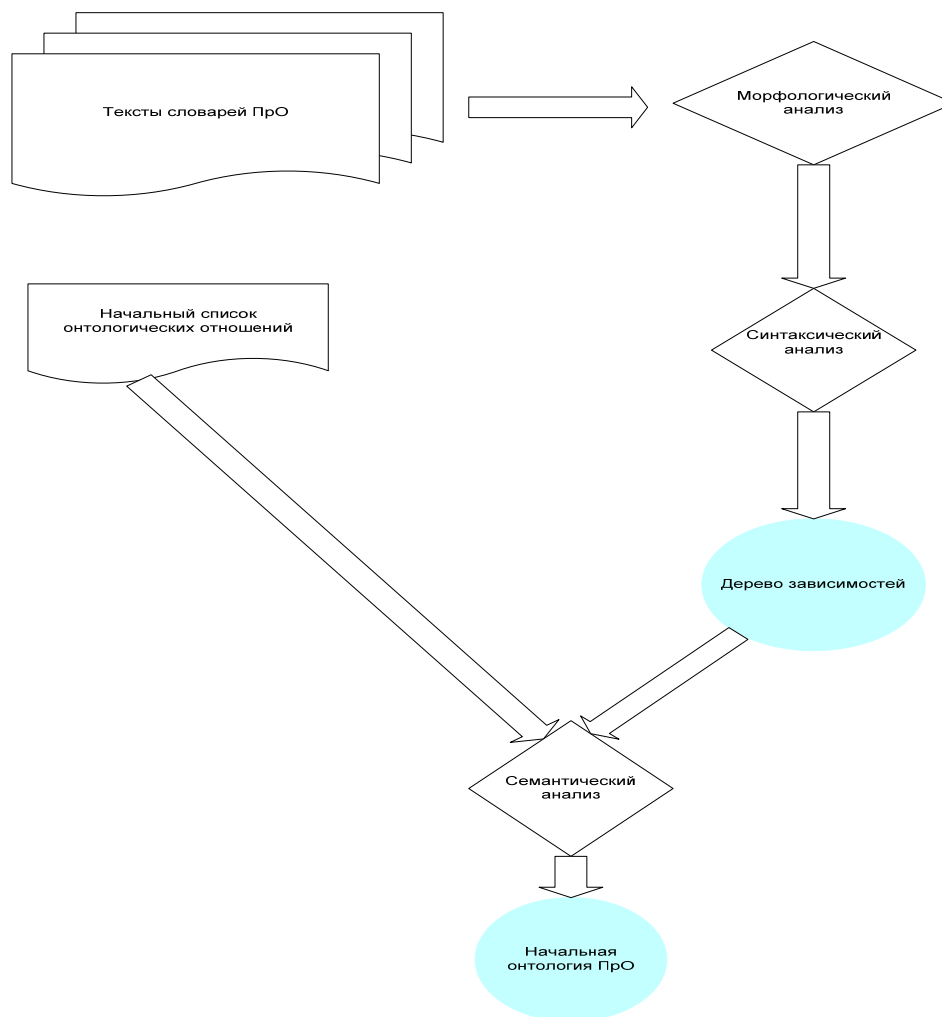


Рис.2. Общая схема первого этапа алгоритма

### Пример использования алгоритма

Рассмотрим в качестве примера процесс построения онтологии ПрО на примере онтологии «Мебель». Вначале построим список терминов, например {стол, стул, комод, шкаф}. Понятие «стол» в данном случае определяется как совокупность таких атрибутов, как высота, ширина, глубина, наличие дополнительных элементов.

Пусть в тексте выделено предложение «Стол угловой компьютерный с надставкой для работы дома "Домашний офис" для работы дома или в офисах переоборудованных из квартир». Понятие «стол» в данном случае определяется как совокупность таких атрибутов, как высота, ширина, глубина, наличие дополнительных элементов. Если в корпусе текстов встречается словосочетание «Стол компьютерный», добавляется класс «\*».

В результате анализа предложения создаются следующие классы

СТОЛ = {{угловой, \*}, (компьютерный, обеденный, офисный)}.

Далее формируются номера классов:

ID {{угловой, компьютерный }} = 1

Назначение = {компьютер, обед}

Дополнительные элементы = {надставки}

Надставки = {для работы дома "Домашний офис"}

Предлог «с» определяет отношение принадлежности.

Затем строка «Размер 105x85x194Нсм» преобразуется в атрибуты ширина, глубина, высота объекта с ID.

Далее, если в тексте встретится объект «Стол угловой офисный», добавляем к множеству стол один элемент ID {{угловой, офисный }} = 2;

Представим множество подклассов класса стол в виде пространства, в данном случае двухмерного. Координатой класса (угловой, компьютерный) теперь будет ID [1][1], а класса (угловой, офисный) = ID [1][3].

В случае, когда в корпусе текстов не найдены словосочетания, например, «Стол офисный» и «Стол компьютерный», размерность первой координаты можно уменьшить с целью уменьшения размерности матрицы ID подклассов.

Теперь для того, чтобы найти в онтологии экземпляр «Стол угловой офисный» достаточно среди множества ID найти элемент с координатой ID [1][3], причем поиск будет выполняться независимо от порядка ввода слов «Стол угловой офисный» или «Стол офисный угловой».

---

### Использование СЛО и онтологии ПрО в процессе лексического анализа.

---

Обе онтологии, построенные в результате работы алгоритма, могут использоваться для решения задач лексического анализа в тех случаях, когда необходима информация о семантике слов. Среди таких задач следующие:

- - определение денотата слова. Например «Во дворе растет каштан. Листья дерева пожелтели». В этом случае в последних двух предложениях надо найти объект  $x$  класса «дерево»
- - восстановление эллипсисов. Например «В магазинах появилась новая книга. Обложка ярко оформлена». В этом случае необходимо установить, что объект *книга* в первом предложении содержит функциональную часть *обложка*, о которой идет речь во втором предложении.
- - разрешение синтаксической неоднозначности. Например «Замки для дверей с несколькими цилиндрами». В данном случае возможны два варианта построения дерева зависимостей: «замки с несколькими цилиндрами» и «двери с несколькими цилиндрами». Поскольку в онтологии ПрО будет существовать класс «замки с цилиндрами», будет выбран правильный вариант.

---

### Выводы

---

Суть предложенной концепции – чтобы автоматизировать построение онтологии предметной области, предлагается использовать специализированную онтологию, имеющую фиксированную структуру и связывающую понятия этой ПрО с определенными лингвистическими единицами – лексемами того или иного ЕЯ. Автоматизируя создание онтологий ПрО, можно достичь интероперабельного использования знаний в современных распределенных знание-ориентированных приложениях. Автоматизация создания онтологии достигается за счет: а) формализованного описания понятий ПрО в виде словарных статей; б) отказа от формализованного описания большого числа отношений. Описываются только отношения, необходимые в данной ПрО.

Полученная онтология позволяет 1) находить информацию в онтологии независимо от формулировки запроса; 2) разрешать различные случаи неоднозначности в текстах; 3) находить денотат для каждого объекта текста. В результате проведенных исследований получен список синтаксических структур,

соответствующих определениям в ЕЯ-текстах. Предложенный метод дает хорошие результаты в том случае, когда имеется большое количество документов при небольшом разнообразии синтаксических конструкций.

Следующим шагом в разработке предложенного метода может быть создание алгоритма распознавания и семантической интерпретации поименованных сущностей в ЕЯ-текстах. Описание процессов в ПРО в виде выполнения определенных операций позволит реализовать цепочку «входное сообщение – система знаний – реакция» [0].

---

## Литература

1. Палагин А.В., Петренко Н.Г. [Системно-онтологический анализ предметной области](#) // УСИМ. – 2009. – № 4. – с.3-14.
2. Величко В., Волошин П., Свитла С. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем. – [www.foibg.com/ibs\\_isc/ibs-15/ibs-15.pdf](http://www.foibg.com/ibs_isc/ibs-15/ibs-15.pdf)
3. Палагин О.В., Свитла 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.
4. Gladun A., Rogushina J. Use of Semantic Web technologies in design of informational retrieval systems // Building and Environment, 2009 Nova Scientific Publishing, New-York, USA. – P.89-103.
5. Gladun A., Rogushina J. Garcia-Sanchez F., Martinez-Bejar R., Fernandez-Breis J.T. An application of intelligent techniques and Semantic Web technologies in e-learning environments // Expert Systems with Applications, An International Journal, 2009, V.36. – P.1922-1931.
6. <http://www.neon-toolkit.org/wiki/1.x/Text2Onto>
7. Musen M. Domain Ontologies in Software Engineering: Use of Protege with the EON Architecture // Methods of Inform. in Medicine, 1998. - P.540-550.
8. IDEF5 Method Report // <http://www.idef.com/pdf/Idef5.pdf>
9. Добров Б.В., Лукашевич Н.В., Невзорова О.А., Федунев Б.Е. Методы и средства автоматизированного проектирования прикладной онтологии // Известия РАН. Теория и системы управления. М.: 2004. № 2. – С.58-68.
10. <http://www.ontologyportal.org>
11. Розробка методів та засобів онтолого-лінгвістичного аналізу природно-мовних об'єктів./ М.Г.Петренко, О.В. Палагін, В.Ю. Величко. С.Л. Кривий - Київ -2009. -38 с. (препр., Інститут кібенетиким ім. В. М. Глушкова.)
12. Отношения в RussNet-e.// [http://project.phil.pu.ru/RussNet/relations\\_ru.shtml](http://project.phil.pu.ru/RussNet/relations_ru.shtml)
13. Виталий Величко, Павел Волошин, Светлана Свитла. АВТОМАТИЗИРОВАННОЕ СОЗДАНИЕ ТЕЗАУРУСА ТЕРМИНОВ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ ЛОКАЛЬНЫХ ПОИСКОВЫХ СИСТЕМ// International Conference “Knowledge-Dialogue-Solution” KDS-2 2009, Kyiv, Ukraine, October, 2009, [http://www.foibg.com/ibs\\_isc/ibs-15/ibs-15-p03.pdf](http://www.foibg.com/ibs_isc/ibs-15/ibs-15-p03.pdf).
14. Одинцов В.В. Стилистика текста. — Изд. 4-е — М. : ЛКИ, 2007. — 262с.
15. G. Miram. Алгоритмы перевода. Вступительный курс по формализации перевода / Translation Algorithms: Introduction to Translation Formalization- К., Эльга, ника-центр. 2004. – 176 с.
16. Лурия А. Р. Язык и сознание. Под редакцией Е. Д. Хомской. Изд-во Моск. ун-та, 1979, 320 с.
17. <http://slovyk.net>
18. Лесько О. Н., Рогушина Ю. В. Использование онтологий для анализа семантики естественно-языковых текстов. // Проблемы програмування, № 3, 2009. – С. 59-66.
19. Арнольд И.В. Стилистика. Современный английский язык // Флинта Издательство, 2002, 384 с.

---

## Информация об авторах

*Ольга Лесько, Институт программных систем, [oll112233@ukr.net](mailto:oll112233@ukr.net)*

*Юлия Рогушина – с.н.с., к.ф.-м.н., доцент, Институт программных систем НАН Украины, Украина, Киев, 44, пр.Глушкова, 03680; e-mail: [jjj@ukr.net](mailto:jjj@ukr.net)*

*Область исследований: технологии Semantic Web, менеджмент знаний*

## КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВЫБОРЕ РЕДАКТОРА ОНТОЛОГИЙ

Дмитрий Климань, Вера Любченко

**Abstract:** *The known editors of ontologies are analyzed, and their main characteristics are identified. Three decision-making criteria to choose ontology editor are formulated.*

**Keywords:** *ontology, ontology editor*

**ACM Classification Keywords:** *I.2.4 Knowledge Representation Formalisms and Methods*

---

### Введение

Представление данных в форме онтологий обеспечивает возможность их автоматической обработки, а также позволяет наглядно и четко сформулировать знания по выбранной тематике. Обучение, как процесс целенаправленной передачи знаний, является естественной областью для использования онтологий.

В [Любченко, 2009] построена онтологическая модель предметной области учебного курса, в частности построена таксономия понятий, описывающая 5 базовых понятий предметной области, и сформулировано множество аксиом, которые формально ограничивают семантику таксономии понятий на основе объяснений. Построенная модель позволяет стандартизировать интегральное представление контента учебного курса и выполнять агрегирования образовательных объектов на основе их логической структуры и разделения содержания и представления.

Очевидно, что онтологическая модель предметной области полезна только в том случае, когда она может быть повторно использована в процессе разработки новых и анализа существующих учебных курсов. Для облегчения преподавателям работы с моделью следует обеспечить возможность автоматического наполнения шаблонных структур модели содержанием. Для этих целей обычно используют редактор онтологий – приложение, поддерживающее создание онтологий или манипулирование ими, предоставляющее возможности работы с одним или несколькими языками представления онтологий, импорт/экспорт в разные форматы, доступ к библиотекам онтологий, визуализацию, машины вывода, языки запросов, поддержку онтологий верхнего уровня и другое. Актуальной является задача выбора редактора онтологий для автоматизации формирования онтологической модели предметной области учебного курса и определения критериев, влияющих на этот выбор.

Постановка задачи. Необходимо определить основные критерии принятия решений при выборе редактора онтологий в условиях определенной базовой онтологической модели.

---

### Популярные редакторы онтологий

Редакторы онтологий – сравнительно молодая группа программных приложений. Самым старым из них является *Java Ontology Editor – JOE* (<http://www.cse.sc.edu/research/cit/demos/java/joe/joeBeta-jar.html>), последняя версия которого вышла 19 мая 1998 года. Этот редактор был создан в виде Java-апплета и поддерживал минимальные онтологические конструкции, такие как сущности, отношения и атрибуты.

На сегодняшний день редакторы онтологий значительно расширили свои возможности. Наиболее удобные среди них, как правило, являются коммерческими. К таким можно отнести *Be Informed Suite* (<http://www.beinformed.nl>) – программу, разработанную голландской компанией Be Informed, которая содержит удобный графический интерфейс, алгоритмы предположения и экспорт в стандартные форматы. Она была создана в первую очередь для разработки больших приложений на основе онтологий.

Ещё один довольно крупный коммерческий редактор онтологий *LinKFactory* (<http://www.nuance.com/healthcare/products/nlp.asp>) был создан компанией Language and Computing на базе Java-технологий. Редактор имел графический интерфейс, поддержку многопользовательской работы, поддержку нескольких миллионов объектов знаний и минимальное время обработки. После того, как

Language and Computing стала составной частью Nuance Communication Inc., редактор был переименован в Natural Language Processing (NLP) и используется только в сфере медицины.

К узкоспециализированным редакторам онтологии можно отнести редакторы *myWeb* (<http://www.ontologyonline.org/main.html>), *OBO-Edit* (<http://geneontology.sourceforge.net/>) и *OntoBuilder* (<http://iew3.technion.ac.il/OntoBuilder/>). Первые два редактора предназначены только для сферы биологии. Редактор *myWeb* позволяет создавать онтологии в сфере молекулярной биологии, описывающие молекулярные функции и биологические процессы. Он имеет плагин для Firefox – *jOWL* и поддерживает выполнение запросов. *OBO-Edit* был создан в рамках Gene Ontology Consortium, он решает сходные с *myWeb* задачи, но в отличие от *myWeb* распространяется бесплатно с открытым кодом. *OntoBuilder* был разработан для применения в сфере медицины с целью создания словарей данных и реализован в виде веб-сервиса.

Среди свободно распространяемых редакторов онтологий можно выделить онлайн-редакторы, редакторы в виде расширений для Firefox и редакторы в виде отдельных приложений. Бесплатные редакторы, как правило, создаются различными университетами для использования в образовательных целях. Университет Флориды создал *MapTools Ontology Editor* (COE) (<http://coe.ihmc.us/groups/coe/>), написанный на Java и содержащий пакет программных инструментов для построения, совместного использования и просмотра онтологий, использующих Ontology Web Language (OWL). Для концептуальных карт разработан интерфейс, благодаря которому в понятной форме отображаются структура, содержимое и масштаб онтологии.

Университет города Осака, Enegate Co Ltd. и Department of Knowledge Systems (Mizoguchi Laboratory) разработали редактор *Hozo* (<http://www.hozo.jp/>), предназначенный специально для создания тяжеловесных и хорошо продуманных онтологий. Новая версия редактора поддерживает импорт OWL. С каждой новой версией редактора публикуются Hozo API ядра Hozo, что позволяет сторонним разработчикам использовать ядро при создании своих собственных Java приложений.

Похожие редакторы были созданы университетами Манчестера и Мэриленда – *OilEd* (<http://www.cs.man.ac.uk/~horrocks/Publications/download/2001/oiled-dl.pdf>) и *SWOOP* (<http://code.google.com/p/swoop/>), соответственно. Оба редактора являются приложениями, созданными на базе Java-технологии и способными отображать и редактировать OWL онтологии. Единственным отличием является поддержка редактором *OilEd* возможности многопользовательской работы. В настоящее время разработка новых версий редакторов больше не ведётся.

Стэнфордскому университету принадлежит разработка 3 редакторов онтологий – *Protégé* (<http://protege.stanford.edu/>), *Chimaera* (<http://www.ksl.stanford.edu/software/chimaera/>), *Ontolingua* (<http://www.ksl.stanford.edu/software/ontolingua/>). Наиболее известным и успешным является *Protégé*. Этот бесплатный редактор с открытым исходным кодом был создан совместно с университетом Манчестера на базе Java-технологий с использованием набора графических библиотек Swing. Платформа *Protégé* поддерживает 2 различные возможности создания онтологий – с помощью редактора *Protégé-Frames* и с помощью редактора *Protégé-OWL*. Редактор имеет и онлайн-версию под названием *WebProtégé*. Функции *Protégé* постоянно расширяются за счёт создания плагинов сторонними разработчиками. Имеется поддержка возможности экспорта созданных онтологий в различные форматы.

Два других проекта оказались менее успешны. *Chimaera* поддерживает лишь работу с форматами OWL и DARPA Agent Markup Language (DAML). Основная функция данного редактора – объединение нескольких онтологий. Он способен решить задачи реорганизации таксономии и разрешение конфликта имён. *Ontolingua* является веб-сервисом, способным создавать, редактировать и модифицировать онтологии. Сервер поддерживает 150 активных пользователей.

Очень успешным редактором, способным составить конкуренцию *Protégé*, является *TopBraid Composer* ([http://www.topquadrant.com/products/TB\\_Composer.html](http://www.topquadrant.com/products/TB_Composer.html)), созданный компанией TopQuadrant. Полная версия редактора коммерческая, но имеется также и бесплатная версия с меньшим функционалом.

Программа написана на Eclipse, имеется графический интерфейс, возможность загрузки, редактирования и сохранения RDF/XML файлов, создания и выполнения запросов SPARQL (SPARQL Protocol and RDF Query Language), импорта и преобразования в RDF данных из XML, UML, каналов RSS, Atom и реляционных баз данных, создание XML-схемы из RDF/OWL, работа с базами данных RDF, визуализация RDF данных с помощью графиков и схем и многое другое.

Стоит отметить и редактор с открытым исходным кодом KAON (KArlsruhe ONtology) (<http://kaon.semanticweb.org/>), созданный университетом Карлсруэ. KAON является редактором онтологий, ориентированных на бизнес-приложения. Включает в себя полный набор инструментов, позволяющих легко создавать онтологии. Стойкость механизмов KAON основывается на реляционных базах данных.

---

### Ключевые характеристики редакторов онтологий

---

Анализ наиболее популярных редакторов онтологий позволяет заметить, что все они созданы с использованием Java-технологии, что обеспечивает возможность их использования на различных программных платформах.

Кроме этого все редакторы обеспечивают графический интерфейс для редактирования и навигации. Наличие удобного графического интерфейса с набором простых форм для редактирования и простого меню позволяет завоевать популярность у пользователей.

Большинство редакторов поддерживает несколько форматов описания онтологий, что обусловлено отсутствием единого широко используемого формата и оставляет право выбора за пользователем.

Следует отметить, что в последнее время становится обязательным для редакторов поддерживать выполнение логических выводов, а именно выполнять проверку непротиворечивости построенной модели и обеспечивать возможность выполнения запросов.

Положительным фактором является использование в редакторах онтологий реляционных баз данных, использование которых позволяет сделать работу редактора более стойкой, увеличить скорость работы редактора, значительно повысить максимальное количество объектов, с которыми способен работать редактор и уменьшить время восстановления.

Следует отметить тенденцию к обеспечению поддержки многопользовательского режима, что особенно актуально в крупных проектах. Но отсутствие этой возможности нельзя назвать критической.

Основные характеристики проанализированных редакторов приведены в табл.1.

Анализ таблицы позволяет сформулировать три критерия принятия решений при выборе редактора онтологий:

1. Если определена базовая онтологическая модель, то следует использовать специализированные критерии.
2. Если необходимо обеспечить поддержку и возможность выполнения импорта/ экспорта онтологий в различные форматы, то в качестве альтернатив следует рассматривать универсальные редакторы. Для специализированных редакторов характерна поддержка ограниченного количества форматов, как правило, формата расширяемого языка разметки (XML) и формата языка онтологии для Интернет (OWL-RDF).
3. Если необходимо обеспечить возможность выполнения запросов к онтологической базе знаний, то следует использовать специализированный редактор. В силу того, что специализированные редакторы разрабатываются на основе базовой онтологической модели, существует возможность обеспечить эффективный механизм выполнения запросов.

Таблица 1

Редактор онтологий	Поддерживаемые форматы	Наличие базовой модели	Проверка непротиворечивости	Выполнение запросов	Многопользовательский режим	Открытый код
Универсальные редакторы						
Be Informed Suite	XML; RDF; OWL; XTM; LKIF	-	-	+	-	-
СmapTools Ontology Editor	XML; RDF; OWL	-	+	-	-	-
Hozo	XML; RDF; OWL; CSV; DAML	-	+	+	-	-
KAON	RDF(S); OWL	-	+	+	-	+
OilEd	DAML+OIL/OWL; RDFS; SHIQ	-	-	-	+	-
Ontolingua with Chimaera	DAML+OIL; KIF; OKBC; Loom; Prolog; Ontolingua; CLIPS	-	-	-	-	-
Protégé	RDF(S) ; OWL; XML Schema; RDB schema	-	-	+	-	+
SWOOP	OWL	-	+	+	-	+
TopBraid Composer	XML; RDF; OWL; RDBM; RSS/Atom; UML; HTML	-	+	+	-	-
Специализированные редакторы						
LinKFactory	XML; RDF(S); DAML+ OIL/OWL	+	+	+	+	-
myWeb	OWL-RDFS	+	-	+	-	-
OBO-Edit	XML; RDF	+	-	+	-	+
OntoBuilder	XML; RDF; HTML	+	-	+	-	-

### Заключение

Для определения критериев принятия решений при выборе редактора онтологий в условиях существующей базовой онтологической модели были проанализированы характеристики наиболее популярных редакторов онтологий. Анализ показал, что все редакторы можно разбить на две группы: универсальные редакторы, которые не привязаны к конкретной предметной области и конкретному формату представления, и специализированные редакторы, которые разрабатываются на основе существующей базовой онтологической модели. Таким образом, для работы с онтологической моделью предметной области учебного курса необходимо использовать специализированный редактор онтологий.



При разработке этого редактора не стоит обеспечивать поддержку импорта/экспорта в различные форматы. Если учесть, что результаты работы этого редактора онтологий должны использоваться различными информационными системами, точный состав которых и требования к формату входных/выходных данных неизвестны, то данное утверждение является обоснованным. При уточнении требований проблему несовместимости форматов можно будет решить программными средствами, путем введения мостов или оберток.

Следует обратить внимание на то, что при разработке редактора онтологической модели предметной области учебного курса в обязательном порядке должен быть обеспечен механизм выполнения запросов к модели. Кроме того, желательно обеспечить поддержку проверки непротиворечивости формируемой модели.

---

## Литература

[Любченко, 2009] Любченко В.В. Онтологічна модель знань про предметні області навчальних курсів // IX міжнародна наукова конференція ім. Т.А.Таран «Інтелектуальний аналіз інформації ІАІІ-2009», Київ, 19-22 мая 2009 г.: сб.тр. — К.: Просвета, 2009. — С. 269 — 275.

---

## Информация об авторах



**Дмитрий Климань** – бакалавр, Одесский национальный политехнический университет, пр. Шевченко, 1, Одесса, 65044, Украина; e-mail: 2842150@gmail.com

Научные интересы: системы онлайн-обучения, онтологии



**Вера Любченко** – доцент, Одесский национальный политехнический университет, пр. Шевченко, 1, Одесса, 65044, Украина; e-mail: vira.lyubchenko@gmail.com

Научные интересы: системы электронного обучения, системы поддержки принятия решений, формальные методы для разработки программного обеспечения

## INTEGRATION OF FINANCIAL DOMAIN KNOWLEDGE ON BASE OF SEMANTIC WEB TECHNOLOGIES

Anatoly Gladun, Julia Rogushina, Rodrigo Martínez-Béjar, Francisco García-Sanchez and Rafael Valencia-García

**Abstract:** *The increasingly huge volume of financial information found in a number of heterogeneous business sources is characterized by unstructured content, disparate data models and implicit knowledge. As Semantic Web Technologies mature, they provide a consistent and reliable basis to summon financial knowledge properly to the end user. In this paper, we present SEFSS, a semantically enhanced financial search engine empowered by semi-structured crawling, inference-driven and ontology population strategies bypassing the present state-of-the-art technology caveats and shortcomings.*

**Keywords:** *Semantic Web Technologies, financial knowledge, ontology, Web-services, data crawling*

**ACM Classification Keywords:** *1.2.4 Knowledge Representation Formalisms and Methods - Semantic networks, H.3.3 Information Search and Retrieval - Retrieval models*

---

### Introduction

---

The increasing availability of financial data collected by banks and other organizations attached to specific applications, pushes the need of real time access very urgent. The optimal solution would allow accessing financial data the way we currently browse unstructured information on the web with popular search engines. Unfortunately, this is not yet possible, since data, residing in thousand incompatible formats, not limited to different data technologies or supporting software, is also incompatible at semantic level. The problem, referred as semantic non-interoperability, is due to differences as the “world-view” level, since related data fields across repository possibly contain data with different meaning, coding scheme or format. Considering this issue on a much broader scale, it applies to thousands of data structures within thousands of databases and messaging formats, spread across the planet. All Semantically-Empowered Financial Search Systems (SEFSS) aims to address these issues by adopting a state-of-the-art semantic knowledge management approach, by extracting meaning from financial data, and unleashing hidden relationships between related but incompatible datasets.

SEFSS will provide access to such relationships by developing a semantically-empowered financial search engine: financial data, crawled from structured and unstructured information sources both publicly available on the Internet and provided by private corporate information sources, will be modeled and semantically annotated. The process will be mostly automated, adopting a supervised information extraction approach based on natural language processing, in order to properly index information fed by harvesting the available sources and extract significant metadata. An improved access level will be provided by exploiting the potential of semantic classification and inference on a domain ontology developed within the project. The lightweight RDF format, used for representing metadata and relationships, will ease this task, especially on the inference and retrieval side. On the language side, although only the handling of English language information will be supported, a particular care will be given to make the system upgradeable to support other EU languages in an easy way.

---

### Concept and objectives

---

The need to manage financial data has been coming into increasingly sharp focus for some time. Years ago, these data sat in silos attached to specific applications in banks and financial companies. Then the Web came into the arena, bringing the hurly-burly of data becoming available across applications, departments and entities in general. However, throughout these developments, a particular underlying problem has remained unsolved: data reside in thousands of incompatible formats and cannot be systematically managed, integrated, unified or cleansed. To make matters worse, this incompatibility is not limited to the use of different data technologies or to

the multiple different “flavors” of each technology (for example, the different relational databases in existence), but also because of its incompatibility in terms of semantics.

The main goal of this work is to develop a semantically empowered financial search engine platform. This SEFSS is designed to gather financial data from very different sources and store them semantically annotated in a repository. This information gathering will occur from both public information sources such as the Internet and from corporate private information sources. Users will then be provided with different services for accessing the data. These services will take advantage of the machine-readable semantic annotations of the financial information in order to provide more sophisticated high-quality functionality to the system’s users.

SEFSS includes two functional modules:

1) *Analyst Information Assistant*

The mission of the Analyst Information Assistant Module will be to gather information from such resources about financial information, including financial information web sites, economic opinion sources, financial information aggregators, blogs and any related Web source.

2) *Trader Information Decision*

Searching and integrating data from various sources has become a fundamental issue in financial research, particularly in those fields where massive data gathering is faced. The reason for this is that the need for information integration in such fields is critical, preserving by all means the semantics inherent to the different data sources and formats. In terms of the financial investment and trading domain, such integration would permit to organize properly data fostering the analysis and access of such information to accomplish critical tasks such as investment influence, trading history and analysis of investing trends at a particular time. The mission of this module will be to bring these techniques to its full potential.

It is also important to highlight two major issues regarding the platform functionality. First of all, the system will be capable of accessing both (semi-)structured and natural language-based data sources, thus embracing most of the data available on the Internet. Secondly, although at a first stage SEFSS will focus exclusively on financial data sources in English, the system to be developed will be designed so that it is straightforward to upgrade it to handle other languages like Spanish or German.

Using the vast consortium expertise in semantic, knowledge acquisition and agents technologies, SEFSS concentrates on building an innovative solution for search and management of financial data. To achieve such an ambitious goal a set of functionalities need to be developed starting with *extraction of relevant meaning from structured and unstructured information*, and ending with *information search*.

In development of SEFSS we use a lot of IT and artificial intelligence results:

- Advanced knowledge management systems for information-bound organizations and communities, capable of extracting actionable meaning from structured and unstructured information and social interaction patterns, and of making it available for activities ranging from information search through conceptual mapping to decision making.
- Semantic foundations: probabilistic, temporal and modal modeling and approximate reasoning through objective-driven research moving beyond current formalisms.
- Service architectures, platforms, technologies, methods and tools that enable context-awareness and discovery, advertising, personalization and dynamic composition of services.

---

## **Semantic Web**

---

The project described in this proposal involves the development of a software application based on four cornerstone technologies: the Semantic Web and ontologies, agent technology, logical reasoning and inference, and knowledge acquisition from texts. Next, these technologies state-of-the-art is described and the advances this project aims to bring about are detailed.

### **a) Semantic Web and Ontologies (data crawling and storing)**

The World Wide Web (WWW), also called “the Web”, was invented in 1989 by Tim Berners-Lee and his colleagues at CERN and it changed the way people gather and access information. Nowadays, the Web is an ever-growing huge data repository. As a consequence, a major bottleneck has emerged when trying to exploit the information represented in the Web, namely, how to find a specific piece of information we may be interested in., ontologies are the backbone technology of Semantic Web [Semantic Web]. Ontologies provide a common vocabulary of an area and define – with different levels of formality - the meaning of the terms and the relations between them [Fensel, 2002]. OWL (Web Ontology Language) (Web Ontology Working Group) is the *de facto* Semantic Web standard ontology language [Maedche, 2001]. A major problem for the success of the Semantic Web vision is the difficulty associated to the semantic annotation of the information already available on the Web. A key contribution of this work is the development of Web crawlers that, in a (semi-)automated way, are able to create RDF [RDF model, 1999] triples from structured, semi-structured and non-structured data on the Web.

Financial Data Crawling and Storage is based on a software component for the crawling, analysis and storage of financial data represented in several formats, extracting RDF metadata records to be used for search and retrieval, according to a Financial Ontology that serves as a unifying data model. This Financial Ontology is found with the requirements of the case studies and a number of alternatives are envisaged to make it consistent and self-contained.

### **b) Agent Technology**

The intelligent agents and multiagent systems area has received ever-increasing attention by researchers over the last few years. *Agents* are the computer systems in charge of carrying out this task. Agents can be useful as stand-alone entities that are delegated particular tasks on behalf of a user. However, in the majority of cases agents exist in environments that contain other agents, constituting multiagent systems (MASs) [Gladun, 2005]. A MAS can be seen as a system consisting of a group of agents that can potentially interact with each other [Gladun, 2006].

Nowadays, the agent community is facing the problem of integrating agent technology with Web services. The SEFSS project will lead to the development of a reference model by means of which both intelligent agents and Web Services can seamlessly communicate to each other, so being able to cooperate in a global environment. We use MAS paradigm because financial domain is too complex to be processed by alone agent [Gladun,2006].

### **c) Reasoning, Inference Engine**

Ontology reasoning is a research area intensively investigated in the recent years. Most of the techniques and inference engines developed for Semantic Web data are focusing either on reasoning over instances of an ontology with rules support (e.g. Rule based approaches) or on reasoning over ontology schemas (DL reasoning) [Baader, 2003]. The examples of such tools are Pellet [Pellet, 2009], FaCT++ [FaCT++ , 2009] - is the new generation of the well-known FaCT OWL-DL reasoner, KAON2 [KAON2, 2009] - is an infrastructure for managing OWL-DL, SWRL, and F-Logic ontologies.

Reasoning over instances of an ontology, for example, can derive a certain value for an attribute applied to an object. These inference services are the equivalent of SQL query engines for databases, but they provide stronger support (for example, handling of recursive rules). Reasoning over concepts of an ontology, for example, can automatically derive the correct hierarchical location of a new concept in a given concept hierarchy. Nowadays also the integration of rule and DL based reasoning approaches gathered a lot of attention. However most of the approaches are considering rather static information, which does not apply to a dynamic domain such as the financial sector. In SEFSS we plan to build reasoning techniques which deal in a scalable way with the dynamism of the financial sector.

### **d) Knowledge Acquisition for Ontologies from Natural Language Texts**

There are different techniques for ontology learning: symbolic techniques, statistical techniques and machine learning techniques. There are two main trends in ontology building: (i) manual ontology building, and (ii) (semi-) automatic ontology building (i.e. ontology learning). Ontology population is the process through which a given

ontology is populated with instances. SEFSS needs to import data from a wide range of channels, being it structured or unstructured [Valencia-Garcia, 2004]. We plan to design component for the crawling, analysis and storage of financial data represented in several formats, extracting RDF metadata records to be used for search and retrieval [Rogushina, 2006].

This tool gets the reference ontology and natural language free texts as inputs and, using a set of NLP tools, it obtains a populated ontology. The ontology population process is comprised of two main phases: NLP phase and Population Phase. The scenario could be as follows (see Fig. 1):

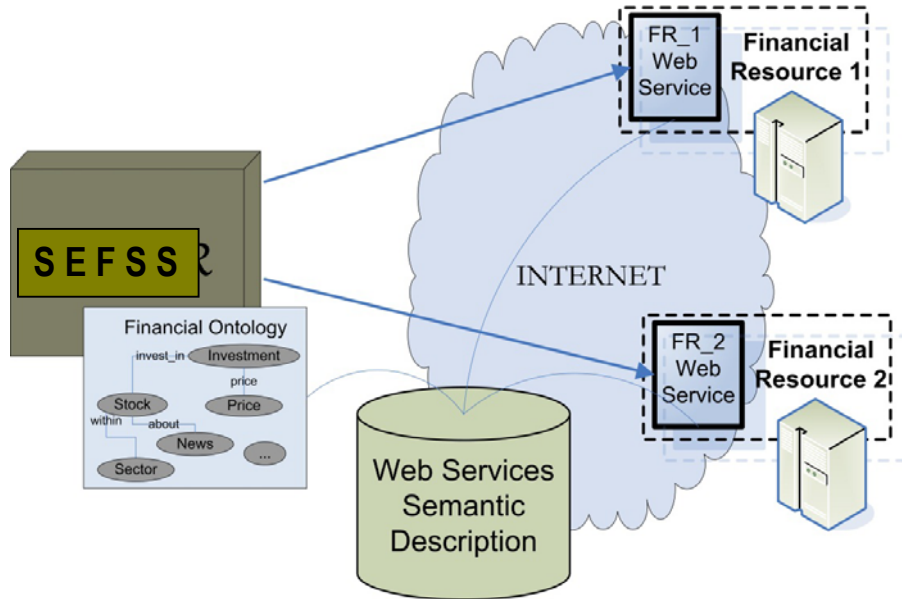


Fig. 1. Crawling the Web for financial information.

The ontology population methodology, based on existing research results from the University of Murcia, will be used for implementing an interface for assisted or supervised processing of free text information, and significant effort will be put to make the process as automated as possible. This methodology gets a reference ontology and natural language free texts as inputs and using a set of NLP tools obtain a populated ontology (see Fig. 2). This ontology must be verified by an expert.

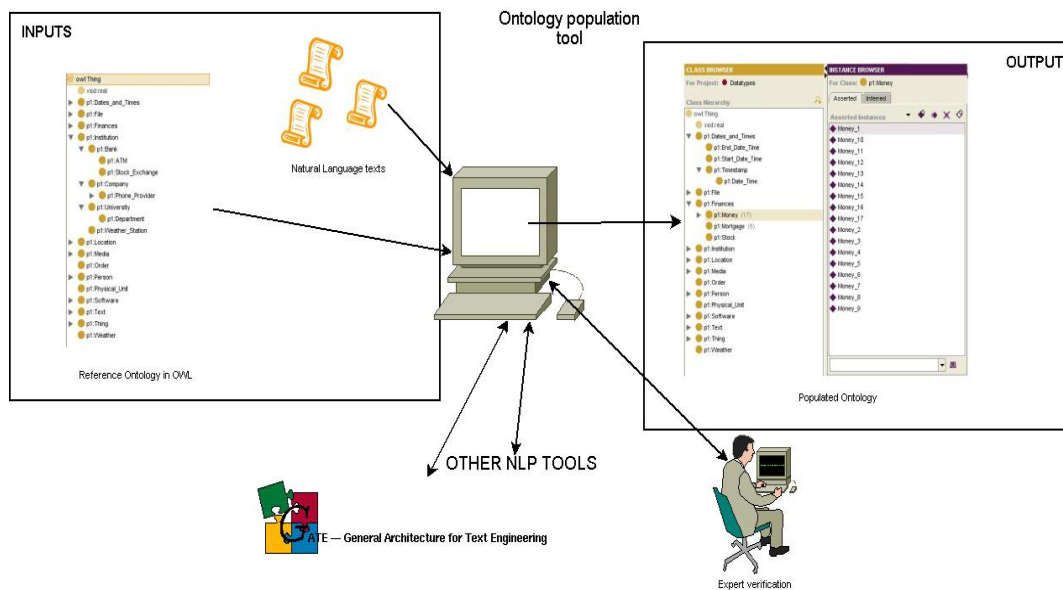


Fig. 2. Ontology Population Tool.

---

## Related Work

---

Web 2.0 is unleashing a number of possibilities that, combined with the Semantic Technologies, could result in significant success. Since the work on improving search results spans over and binds together a number of research initiatives, in this section we briefly describe related work.

Searching has been subject of intensive research but a more concrete survey on filtering search results and optimizing results yields also a remarkable amount of efforts. Following research successfully implemented in the Google search engine, a number of search variants related to the work presented have been explored such as using faceted search, including its application to multimedia faceted metadata for image search and browsing or navigating RDF data.

Wherever possible, research results will be exploited for the internal development and support of new products and services. These products and services will lead to a competitive advantage of the participating organizations and will create a substantial benefit for the targeted users. In order for the exploitation to be effective, an integrated approach will be necessary, combining experience and expertise from the development department and solution management, and the involvement of a user base [represented by the consortium partners - if there are no user partners: which will be involved through dedicated surveys]. Integral part of the exploitation approach is the identification of a [one or more] use case which will serve as the validation point throughout the project - from customer requirements, to building the demonstrator up to the evaluation of the results.

The project will create exploitation plans for the results of individual participants and for the consortium as a whole in order to ensure that the developed technologies have a significant impact in the market and that they do not remain as theoretical developments. The SEFSS project will include a specific component and a work package devoted to exploitation issues. The major results of the exploitation work package are the Market Analysis and the Exploitation Plan. As part of the management structure, an Exploitation Management Board will ensure the successful execution of the exploitation work package and will ensure that the other work packages take into account results from the exploitation work package.

The industrial partners within the project have clear exploitation routes for the technology in a number of ways, and their individual exploitation intentions will be stated in the exploitation plan. The industrial partners are all experienced in generating and protecting intellectual property. In some cases software, notably certain software modules generated by the academic partners, may be open-source, which will further encourage the take-up of the technology.

The SEFSS exploitation strategy consists in outline of:

- Tracking important commercial and technical developments in the Semantic Web.
- Analyzing more deeply and in detail of the financial sector.
- Identifying the results of SEFSS that are exploitable: semantically-empowered financial search engine platform.
- Identifying the appropriate distribution channels to exploit the different results of the project.
- Analyzing the impact of the financial search engine platform in terms of quantitative and qualitative impact, when targeting the deployment and integration of the platform in financial entities. This evaluation will be based on real-world implementations.

Different activities will be sought to represent the project actively and establish liaisons with other relevant projects, standard organizations, and institutions that can be of benefit for the project. The different partners will establish contacts with other companies outside the consortium and preparing the market for the technology adoption. These derived products, contacts, potential users and exploitation plans will be documented in a final exploitation plan report.

Coming up, details on the intentions of the project participants in the dissemination activities are:

Metaware has a clear exploitation path for foreseen project results, as a way to add more powerful and insightful technologies and approaches into the existing Business Intelligence products and solutions. The direct peering with top financial and public bodies, thanks to a strong background in the financial services, may also allow providing interesting opportunities for further extension of the results, and technological transfer.

---

### Conclusions and discussion

Conventional wisdom holds that new Semantic Technologies promise a powerful paradigm for solving integration problems with current state-of-the-art IT infrastructure such as financial search engines and applications. However, there has been not significant progress in terms of real developments, particularly because of the Semantic Technologies problem: Web information or data source providers would always request for a good excuse or reason, a good application or benefit from providing metadata. However, if the metadata is not generated, no application or value-added functionality can be achieved. But the explosive growth of a number of structured metadata formats in blogs, wikis, social networking sites and online communities has transformed the Web in recent years. Mainstream media has taken notice of the so-called Web 2.0 revolution and business success stories have gathered stream.

These technologies are blooming overnight and providing “metadata farms” whose potential can be unleashed by Semantic applications, which will gain momentum towards a Web generation in which Semantic and Web 2.0 technologies will end up meeting.

Our future work focuses on tuning and optimizing that crossway and enabling the transition from research and academic prototypes to practice and from standards to deployment. Tool vendors and manufacturers are reluctant to implement products until they see a market forming, but we envisage the market needs as a powerful driving force to make these solutions mature and business oriented.

---

### Bibliography

- [Semantic Web] Semantic Web Challenge. - <http://challenge.semanticweb.org/>.
- [Fensel , 2002] Fensel, D.: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer, Heidelberg (2002).
- [Maedche , 2001] Maedche, A., Staab, S.: Ontology learning for the semantic web. IEEE Intelligent Systems 16(2), 72–79 (2001).
- [RDF Model, 1999] RDF Model and Syntax Specification. W3C Proposed Recommendation. - January 1999. - <http://www.w3.org/TR/PR-rdf-syntax>.
- [Gladun, 2005] Anatoly Gladun and Julia Rogushina: Ontologies as a Perspective Direction of Intellectualization of Informational Retrieval in Multiagent Systems of E-commerce // The Proceedings of XI-th Intern. Conf. "KDS'2005", Varna, Bulgaria, pp.112-120.
- [Gladun , 2006] Gladun A. , Rogushina J. Ontological Approach to Domain Knowledge Representation for Informational Retrieval in Multiagent Systems // International Journal "Information Theories & Applications", V.13, N.4, 2006. – P.354-362.
- [Baader , 2003] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook. Cambridge University Press, Cambridge (2003).
- [Pellet:, 2009] Pellet: OWL 2 Reasoner for Java.- <http://clarkparsia.com/pellet> .
- [FaCT++, 2009] FaCT++ - <http://owl.man.ac.uk/factplusplus/> .
- [KAON2, 2009] KAON2 - <http://kaon2.semanticweb.org/>
- [Valencia-Garcia, 2004] Valencia-Garcia, R., Ruiz-Sanchez, J.M., Vicente, P.J.V., Fernandez-Breis, J.T., Martinez-Bejar, R.: An incremental approach for discovering medical knowledge from texts. Expert Syst. Appl. 26(3), 291–299 (2004)
- [Rogushina , 2006] Rogushina J., Gladun A.: "Semantic Search of Internet Information Resources on Base of Ontologies and Multilingual Thesauruses" // International Journal «Information Theories and Applications», vol.14, 2006.-P.117-129.

---

**Authors' Information**

---



**Dr. Anatoly Gladun** – senior scientific researcher, Associate Professor, International Research and Training Centre of Information Technologies and Systems, National Academy of Sciences and Ministry of Education of Ukraine, 44 Glushkov Pr., Kiev, 03680, Ukraine ; e-mail: [glanat@yahoo.com](mailto:glanat@yahoo.com) .

*Major Fields of Scientific Research: Intellectualization of computer networks, Intelligent Software Agents and multiagent systems, Semantic Web technologies, Information Retrieval.*



**Dr. Julia Rogushina** – senior scientific researcher, Associate Professor, Institute of Software Systems of National Academy of Sciences of Ukraine, 44 Glushkov Pr., Kiev, 03680, Ukraine; e-mail: [jjj\\_@ukr.net](mailto:jjj_@ukr.net)

*Fields of Scientific Research: Semantic Web technologies, knowledge management, data mining. knowledge acquisition.*

**Dr. Rodrigo Martínez-Béjar** - Full Professor, Departamento de Informatica y Sistemas, Univeridad de Murcia, Spain, [rodrigo@um.es](mailto:rodrigo@um.es) .

*Major Fields of Scientific Research: General theoretical information research, Multi-dimensional information systems.*

**Francisco Garcia-Sanchez** - , Departamento de Informatica y Sistemas, Univeridad de Murcia, Spain, [fgarcia@um.es](mailto:fgarcia@um.es) .

*Major Fields of Scientific Research:*

**Rafael Valencia-Garcia** -, Departamento de Informatica y Sistemas, Univeridad de Murcia, Spain, [valencia@um.es](mailto:valencia@um.es) .

*Major Fields of Scientific Research:*



## A COMPREHENSIVE APPROACH AND USER-ORIENTED POLICY FOR MANAGING RISK AND SAFETY

**Kristian Milenov, Krassimir Markov, Venko Bozhilov, Radko Radkov, Pavel Milenov**

**Abstract:** *A possible approach for more flexible and user-oriented risk assessment methodologies and creating infrastructure for exchanging and processing spatial data in Europe is outlined in the paper. Priority is given to setting up a network of servers in different parts of Europe which work in harmony and constitute parts of a unified system, integrated with a network of regional units for risk and security management. Such approach is considered in many respects to be better, more effective and yielding faster and better quality results than the creation of centralized structures – a super-server where primary data are fed by various municipalities, regions and countries, or international super - agency for risk and security.*

**Keywords:** *Infrastructure for spatial data in Europe, user-oriented risk assessment methodologies, network of servers, network of regional units for risk and security.*

**ACM Classification Keywords:** *C.2 COMPUTER-COMMUNICATION NETWORKS, C.2.5 Local and Wide-Area Networks,*

---

### Introduction

If one needs to know what is the time, he or she may ask someone and will know "the local" time. But if one wants to know what time is pointed by all watches in the town or in the country, it is really impossible to solve such problem because of the high number of devices - it is impossible in the same moment to scan and receive all information. During processing one part of watches, the other will show new time. So, one will never know what time is on all watches.

In other words, along with modern technologies appears a lot of information. Often, when we want to collect and understand it, the traditional methods are not enough. They might be too general and they need additional processing and transformation. Therefore we use advanced methods of data processing such as rough sets, genetic algorithms, neural nets or fuzzy sets. Rough sets have application e.g. in problems of excessing data, problems with correct classification or problems with retrieval hidden relations between data. Placing these methods in environment of distributed applications, based on .Net platform and XML Web Services, introduce new possibilities in usage of artificial intelligence methods in data processing systems [Zielosko and Wakulicz-Deja, 2005].

Technological advancement in using intelligent techniques has provided solutions to many applications in diverse engineering disciplines. In application areas such as web mining, image processing, medical, and robotics, just one intelligent data processing technique may be inadequate for handling a task, and a combination or hybrid of intelligent data processing techniques becomes necessary. The sharp increase in activities in the development of innovative intelligent data processing technologies also attracted the interest of many researchers in applying intelligent data processing techniques in other application domains. [Wai Wong et al, 2007]

A briefly classification of the existing methods of data processing by the complexity of results that we want to achieve is given in [Nguyen et al, 1997]:

- In some problems, all we want is one or several numerical values. It may be that we measure some characteristics, or it may be that we know the model, and we want, based on the experimental data, to estimate the parameters of this model. These problems are usually handled by statistical methods.
- In other problems, we want to know a function. We may want to reconstruct an image (brightness as function of coordinates), we may want to filter signal (intensity as a function of time), etc. These methods are usually handled by different regularization techniques.
- Finally, there are even more complicated problems, in which we want to reconstruct a model of an analyzed system. Methods that handle such problems are called **intelligent data processing methods**.

Many of these methods are based on logic programming, a formalism that (successfully) describes complicated logical statements algorithmically, in a kind of programming language terms. Other methods are based on advanced information modeling technologies which use rough sets, genetic algorithms, neural nets, fuzzy sets, etc.

A very important area for applications of the intelligent systems is the European Earth Observation Program (GMES – Global Monitoring for Environment and Security) which provides data useful in a range of issues including climate change and citizen's security. Land, sea and atmosphere – each Earth component is observed through GMES, helping to make our lives safer. The purpose of GMES is to deliver information which corresponds to user needs. The processing and dissemination of this information is carried out within the "GMES service component" [GMES, 2010].

The thematic areas within the GMES service component comprise:

- land, marine and atmosphere information – ensuring systematic monitoring and forecasting the state of the Earth's subsystems at regional and global levels;
- climate change information – helping to monitor the effects of climate change, assessing mitigation measures and contributing to the knowledge base for adaptation policies and investments;
- emergency and security information – providing support in the event of emergencies and humanitarian aid needs, in particular to civil protection authorities, also to produce accurate information on security related aspects (e.g. maritime surveillance, border control, global stability, etc.)

Managing natural resources and biodiversity, adapting to sea level rise, monitoring the chemical composition of our atmosphere: all depend on accurate information delivered in time to make a difference.

The GMES service component depends on Earth observation data, collected from space (satellites), air (airborne instruments, balloons to record stratosphere data, etc.), water (floats, shipboard instruments, etc.) or land (measuring stations, seismographs, etc.). These facilities are called the GMES infrastructure component; non-space based installations in the GMES infrastructure component are generally referred to as "in situ component".

By securing the sustainability of an information infrastructure necessary to produce output information in the form of maps, datasets, reports, targeted alerts, etc., GMES helps people and organizations to take action, make appropriate policy decisions and decide on necessary investments. GMES also represents a great potential for businesses in the services market, which will be able to make use of the data and information it provides according a full an open access principle.

Earth observation-based services already exist in Europe, but they are dispersed at national or regional level and cannot rely on a sustainable observation capacity. With the exception of meteorological services, long-term availability and reliability of information is not guaranteed. This is why, in order to contribute to improve its response to ever growing challenges of global safety and climate change, Europe develops a sustained and reliable Earth observation system of its own [GMES, 2010].

The growing frequency of natural disasters, their serious impacts and huge damage have become to be considered one of the modern challenges of the countries in Europe. Some of the weak points in the European policy and activities in this field are - the lack of prognosis and user-oriented preliminary risk assessment of potential damages in different regions as well as a simple to be used but powerful in data processing information infrastructure. The issue of having a preliminary assessment of the likely damage and ways to mitigate or even avoid it has attained an ever bigger significance but it must be developed in parallel with improving and fostering the transfer of data, information and practices, without losing the local and regional user's priorities.

The experience gained in recent years from various national and international projects aimed at finding solutions to minimize damage from natural disasters and creating workable models of forecasting and assessing potential future damage reveals two components as being major factors for achieving good results: (1) a good methodology and (2) access to the necessary spatial data and information systems. Given that there have already been developed methodologies and simulation models of the respective disaster, what remains an open issue is ensuring the necessary data, in an appropriate format, with the relevant compatibility so that they can be

combined with other data and processed as a package so as to obtain the required synergy effect and results as close as possible to the real ones. The basic approach, presented here is to start with developing a simple and much general model of a possible natural hazard ( an earthquake in this case), using the available information from different sources and open to be upgraded and improved parallel to the increase of the amount of available data sources.

One possible approach for creating an infrastructure to ensure such exchange of spatial data and their processing is to set up a network of servers in different parts of Europe which will work in harmony and constitute parts of a unified system. Such an approach is considered in many respects to be better, more effective and yielding faster and better quality results than the creation of a central server where primary data are fed by various municipalities, regions and countries.

In this connection some key aspects of information security and use of information technology in information systems related to management of Critical Infrastructures, should be considered [Polimirova et al, 2007a]. The main point in the analysis of the threats to the Information technology is the planning and implementation of attacks to information resources in general [Nickolov and Polimirova, 2010] and in particular to the information objects [Polimirova et al, 2007b]. Depending on the environment and client-server architecture different types of protection of information systems used in Critical Infrastructures can be built [Polimirova and Nickolov, 2009], [Polimirova and Nickolov, 2010]. It is necessary to investigate also the issues related to the risk assessment that could be applied to information objects, which are a part of information systems operating Critical Infrastructure [Polimirova, 2008]. This concept closely correspond to the proposal of the European-Mediterranean network of regional units for risk and security management, presented on the GMES Workshop in Sofia, in March, 2010 and included in GMES Workshop Sofia Conclusions [GMES-WS, 2010].

---

### **General overview**

---

In the course of several millennia the human evolution turned the human being from someone who inhabited forests and various natural earth cavities into a geological factor who built first small, and then ever bigger and bigger artificial complexes - buildings grouped in even grander zones: population and industrial agglomerations. The matter moved to be input in these urban agglomerations is commensurate with the quantity expelled by major geological disasters - volcanic eruptions, or moved by tectonic processes - earthquakes.

Irrespective of his incessant effort man cannot yet boast that he has tamed the natural force and the cyclic occurrence of hurricanes, earthquakes, fires which wreak havoc and claim many lives. No one has yet "won a battle" against a natural disaster; one can rather speak of avoiding the worst and mitigating the damage. This is why a "winning war" with nature would need a careful preparation of "the next battle" - an appropriate system for forecasting of possible disaster-induced damage and the strategic planning and preventive measures that go together with it.

The growing frequency of natural disasters, their serious impacts and huge damage have become to be considered one of the modern challenges of the countries in Europe and all over the World. The issue of having a preliminary assessment of the likely damage and ways to mitigate or even avoid it has attained an ever bigger significance.

The experience gained in recent years from various national and international projects aimed at finding solutions to minimize damage from natural disasters and creating workable models of forecasting and assessing potential future damage reveals two components as being major factors for achieving good results: a good methodology and access to the necessary spatial data and information systems. Given that there have already been developed methodologies and simulation models of the respective disaster, what remains an open issue is ensuring the necessary data, in an appropriate format, with the relevant compatibility so that they can be combined with other data and processed as a package so as to obtain the required synergy effect and results as close as possible to the real ones. Thus good methods, simulation models and software are reduced to zero when data sources are not available or they are not accurate, updated and easy for use. The problem is related to the adoption of common rules for data exchange, data storage and formatting, etc., which are the subject matter of European

Directive 2007/02/EC [INSPIRE, 2007] setting the legal framework for the establishment and operation of an Infrastructure for Spatial Information in the European Community (INSPIRE) [INSPIRE TR, 2007], but which also call for the introduction of rules vis-a-vis interoperability of the information systems.

Therefore the presented article is combining research and development topic – as modeling preliminary losses from natural disasters (earthquake in this case), and conceptual topic – as the development of a quality assurance reference data base and regional networking for data storage and processing and for risk and security management.

It is common knowledge that proper measures taken beforehand reduce many times over the damage from a possible natural disaster as well as the resources and money necessary to overcome the consequences. Given that a natural disaster may strike at a particular place but can affect a large geographical area sometimes covering several countries, the work should be aimed at achieving coordinated estimates and measures simultaneously, on the local, as well as on the inter-regional, community or even global, level.

This is why the respective methodologies, spatial databases, simulation models and outputs need to be compatible, aimed to achieve a synergy effect and create a platform of interoperability.

Five levels of management are identified: (1) municipality; (2) country; (3) region; (4) European Union; (5) global level. Halfway levels are possible too, such as unification of policies for risk management in a given area of impact, for example the catchments areas of big rivers. A good example of this would be the creation of an integrated system for prevention against flooding and water pollution along the Danube within the future European Danube Strategy [EU Strategy, 2010]. The European Danube Strategy, set up in October 2008 [Declaration, 2008], recognizes the need for an expansion of coordination between the 10 Danube states – of which six are EU member states. The Danube area has a population of 200 M inhabitants, of whom some 70 M live in regions directly situated on the banks of the river.

In respect of earthquakes, an intermediate form of unification could be sought of policies for risk management in areas affected by known seismic foci or geological areas, e.g. significant faults in contact areas of tectonic plates.

Set out below is an opinion about the necessary exchange, and in certain cases, integration of spatial databases and information systems regarding seismic risk.

---

### **Methodological basis**

---

The methodology for forecasting damage caused by natural disasters, particularly earthquakes, mainly targets the residential urban environment and the public utility and industrial infrastructure related to it, which make up the modern city. The methodology complies with the requirements to yield results that are useful for the end users (local and central powers, business entities, civil protection), as well as to tackle the existing problems with the available spatial data, unreliable information and absence of information.

Why is such a methodology an indispensable element of the effective policy for managing risk and safety?

This is so for a number of reasons, namely:

- Buildings differ in type, mode of construction and age, and thus have different vulnerability to catastrophic seismic impacts, so that the damage they will sustain will be different;
- The modern urban society has never had such surplus of energy, resources and temporal capacity so as to allow the construction of buildings able to withstand the possible maximum seismic impacts;
- Rigorous construction supervision is missing (especially in the time passed); many violations are made on town-planning indicators and the requirements for free spaces and distances between buildings, laid down for safety reasons;
- The policy for sustainable development calls for judicious spending of resources for prevention against seismic impacts, allowing such expenditure only where the alternative to building a new earthquake-proof building or the reinforcing of an old one would yield an immediate positive effect for the public,

either in terms of safety and quality of the living space, or as industrial economic infrastructure, or also as protection of an intangible value of historical or emotional essence;

- Recently we have witnessed (in Bulgaria) a dangerous scheme used by developers who opt for fulfilling their interests by contravening the law and even creating a threat to people's lives and health. In the scheme certain buildings which have public significance, constitute monuments of culture, or are banned from demolition by law, are left in disrepair or else their process of degradation is accelerated artificially, so that they reach the stage of self-destruction and can thus be freely removed and replaced by a new development project which usually exceeds considerably the permitted town-planning parameters. The scheme in question poses a double threat - on the one hand the disintegrating building poses a threat to the members of the public and to the buildings and infrastructure next to it (a fatal accident of this type occurred in Sofia only two years ago), and on the other, the new construction usually fails to comply with the safety criteria and standards and becomes a potential threat during a future earthquake.

Driven by various reasons the modern society (and particularly United Europe) has developed adequate methods for observing a given disaster and for managing economic flows during recovery; however, a single methodology is yet missing for forecasting disaster-induced damage, including in terms of individual risks and topics. This is especially true of places where the data necessary for making the estimates are rarely available in the necessary volume, accuracy and credibility.

In such a case the option is to look for a methodology that is "sustainable" and resilient to data incompleteness, a methodology that is nevertheless applicable, albeit with a reduced accuracy of results. The most important thing in this case is for the methodology to provide a good idea about the areas with increased risk and a reliable general assessment of the likely damage. The methodology needs to be able to be applied both at the local level - a neighborhood or a town, as well as at a greater one - a region or a country. At the next stage partnership is to be sought with experts and structures from neighboring countries to ensure interaction between such methodologies of two countries or more, as well as the exchange of data, models and results.

Such a methodology will prove very useful for risk assessments vis-a-vis urban centers with limited data about housing and public utility infrastructure.

The methodology takes into account the state of the spatial databases and information systems in Bulgaria and the actual difficulties in providing full, detailed and objective information. In this respect the situation in Bulgaria is similar to that in many other EU member states, meaning that the methodology can be successfully applied in the EU, as well as in other geographical regions.

The methodology, developed by a team of experts in different areas: geophysics, economy, town planning, informatics, has three successive layers:

1. risk assessment of housing;
2. risk assessment of public utility and economic infrastructure;
3. assessment of economic damage.

Assessment of the so-called "intangible damage" has not been envisaged, as at this stage it proved impossible to validate in the respective software the connection between intangible or information values and current tangible values and assets of the modern city.

Several versions of the methodology have been set out:

- for assessment of earthquake-induced damage of residential buildings;
- for an overall assessment of the seismic risk;
- for estimation of the economic losses from natural disasters.

These methodologies will be outlined shortly below. Detailed explanation will be given in further publications.

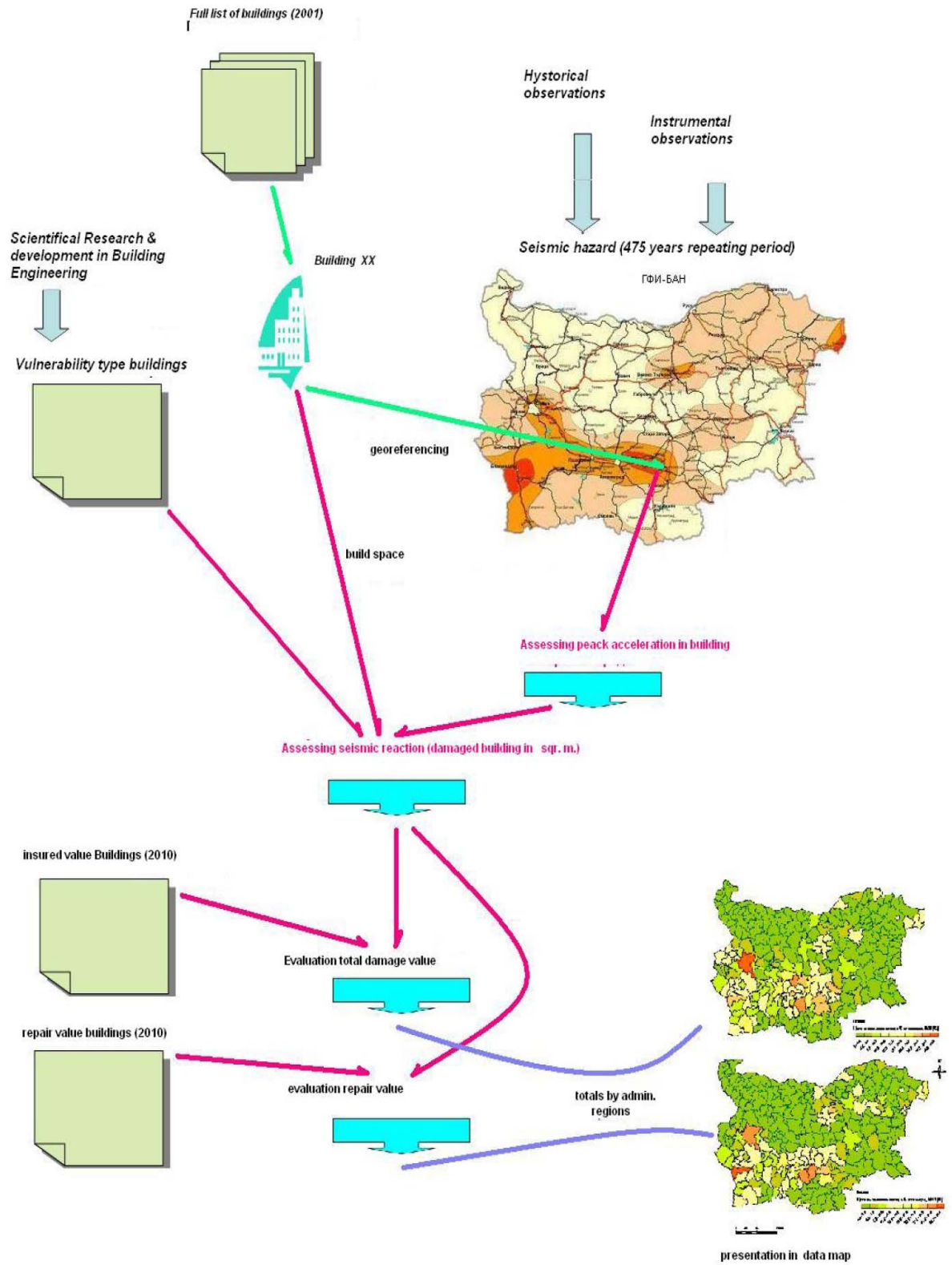


Figure 1. Simplified methodology for seismic damage assessment of dwellings

### Methodology for assessment of earthquake-induced damage of residential buildings at the level of municipality, district, country, region

The methodology has been used on test cases for the Bulgarian cities of Sofia, Ruse, Vraza and for the whole country. First priority was given to results which can be of use for the local and state administration for planning their economic and resources actions as well as for urban planning, building rehabilitation and civil protection.

The methodology is based on experience gained from work under the RISK-UE project in Europe [Mouroux and Brun, 2006] and HAZUS in the United States in the late 1990s and the beginning of the new millennium [HAZUS, 2010]. It starts from:

- availability of data about catastrophic impact (in the region);
- availability of data about housing (in the region);
- availability of a vulnerability function (vis-a-vis the respective housing related to catastrophic impact by type of buildings and type of impact).

To obtain an assessment of the expected damage, systematic modeling is made of the catastrophic impact for each registered housing object, depending on the planned maximum catastrophic impact.

The methodology concerns assessment of the likely damage on residential buildings caused by an expected earthquake with a given magnitude. It is applied on information about housing and the occupants thereof available in the respective National Statistical Service. The methodology is apt to be applied in assessing the value of small to medium damage (for example, up to 20-25% of the area), whereby the recovery value increases non-linearly when the percentage of damage is higher, as costs will be provided for clearance.

The methodology is illustrated by the brief scheme shown on Figure 1.

### Methodology for overall assessment of earthquake damage

The methodology is in the process of being developed; the authors have set themselves the task of extending it (for housing infrastructure) to include the public infrastructure (the urban infrastructure in particular) in the actual situation of missing workable databases about buildings and other urban infrastructure. To this end, partial use is made of the methodology for structuring of the urban infrastructure in the European project RISK-UE (WP3) [Mouroux and Brun, 2006], presented on figure 2.

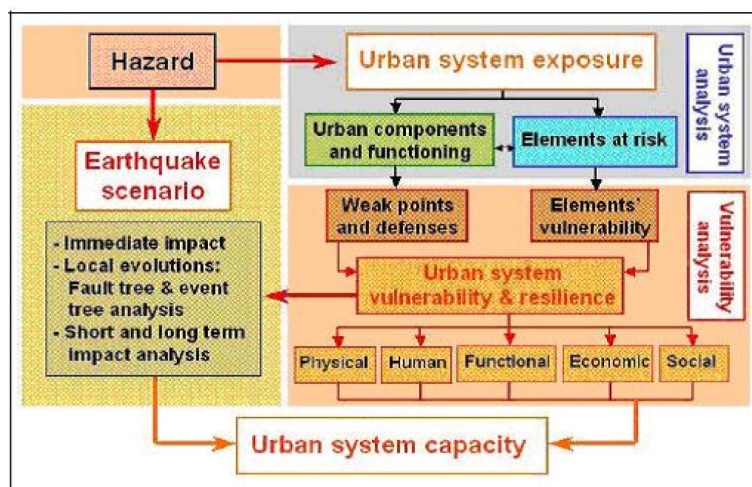


Figure 2. GEMITIS' Seismic Risk Analysis [Masure and Lutoff, 2006]

From the overall approach, the part covering the elements at risk complementing housing is being extracted.

The simplest option would be to apply the basic methodology on sets of infrastructure descriptions. But here lies the main obstacle, in this case - for Bulgaria - the absence of a detailed and comprehensive description of this infrastructure.

---

### **Extrapolation of the housing damages to damages in the public infrastructure**

---

A more complex option would be to arrive at a system for extrapolation of the public housing data to a mechanism for assessment of infrastructure data as a main task in risk assessment.

Solving such a problem is not possible using simple linear extrapolation of housing damage (a purely physical assessment).

In the first place, the purely housing damage - provided that it does not exceed a minimum percentage (e.g., 5%), can be serviced in fully functioning infrastructure (i.e., without, or with only negligible losses in the public infrastructure). Then recovery is seen as part of the normal functioning of society, rather than an emergency activity. This is why losses in the public infrastructure should be considered from a certain threshold and above.

After the initial version, mechanisms can be applied as well for taking into account the impact of losses in the public infrastructure in the sub-scheme of their structuring according to the above mentioned model of WP3-RISK-UE [Mouroux and Brun, 2006]:

- vis-a-vis physical damage (collapsed buildings) of the public infrastructure;
- vis-a-vis functional damage of the public infrastructure (impossibility to function fully);
- vis-a-vis economic damage (from own collapsed facilities and from a diminishing workforce busy with recovery of primary needs);
- vis-a-vis functional and economic damage from newly emerging urgent social&material needs;
- vis-a-vis "social" (non-material historical, marketing and similar) damage as a result of the disaster which diminishes the economic value of the respective area.

Further, mechanisms are applied for transformation of housing damage to damage on the public and economic infrastructure. Some preliminary conditions are accepted as:

- Going beyond a certain percentage of damage (e.g., over 10%) suggests additional costs for rescue, evacuation of population without shelter, setting up of medical aid, security and safety, provision of food for the population remaining in place. Hurricane Katrina, and even more so, the earthquake disaster in Haiti, revealed the considerable shortage of capacity of the security services in the event of a large-scale disaster;
- In the case of more than 10% affected (destroyed) housing, emergency costs are to be provided for temporary housing to shelter the surviving population until recovery of at least 60% of the volume of housing from before the disaster.
- Going beyond a certain percentage of damage (e.g., over 5-10%) suggests non-linear increasing of costs for site clearing. If more than 35-50% of the housing has been affected (has collapsed), from 10% to 40% costs (of the building value) should be provided for site clearing and environmental-friendly processing of waste.
- As a rule, the normal construction activity for a given region is determined as being designed to build/recover no more than 5% of the available housing (i.e., a term of restoration of housing of no less than 20 years). Above this percentage all activities for clearing and restoration constitute additional capacity for construction and repair works which is not available *in situ*.
- More than 35-50% affected (destroyed) housing suggests total destruction and a need to build anew the houses affected by the disaster.
- When the affected (destroyed) housing exceeds 10%, a relevant percentage of destruction of public utility infrastructure should be assumed - schools, hospitals, which could vary from 5% to 15-20% of the value of the housing (depending on the characteristics of the social environment at the place of the disaster).



This should be taken into account, irrespective of the application of specially heightened seismic requirements vis-a-vis such public buildings in certain areas (e.g., in the United States). When more than 40-50% of the housing has been affected, total destruction of the subsurface public utility infrastructure servicing the given housing should be assumed. Respectively, rebuilding of the public utility infrastructure should be provided for.

- In recovery actions in a region heavily hit by a seismic disaster, costs for the transport infrastructure should also be provided for, which need to be higher than the prices of such infrastructure during normal times of construction.
- When more than 15% of the public utility infrastructure has been affected (destroyed), urgent building of temporary replacement public utility infrastructure should be provided for.
- In the event of a seismic disaster (affecting more than 5% of the housing), shortage of construction capacity should be assumed at the place of recovery and a subsequent increase (sometimes multiple) of the economic value of the restoration in relation to the own value of the destroyed infrastructure (both housing and public infrastructure).

### Extrapolation of the housing damages to the economy

Together with the social and public utility infrastructure, a seismic disaster will also damage the economic infrastructure of the given region. In such case, the increase of damage is non-linear and non-proportionate (as set against that of housing) (Figure 3.).

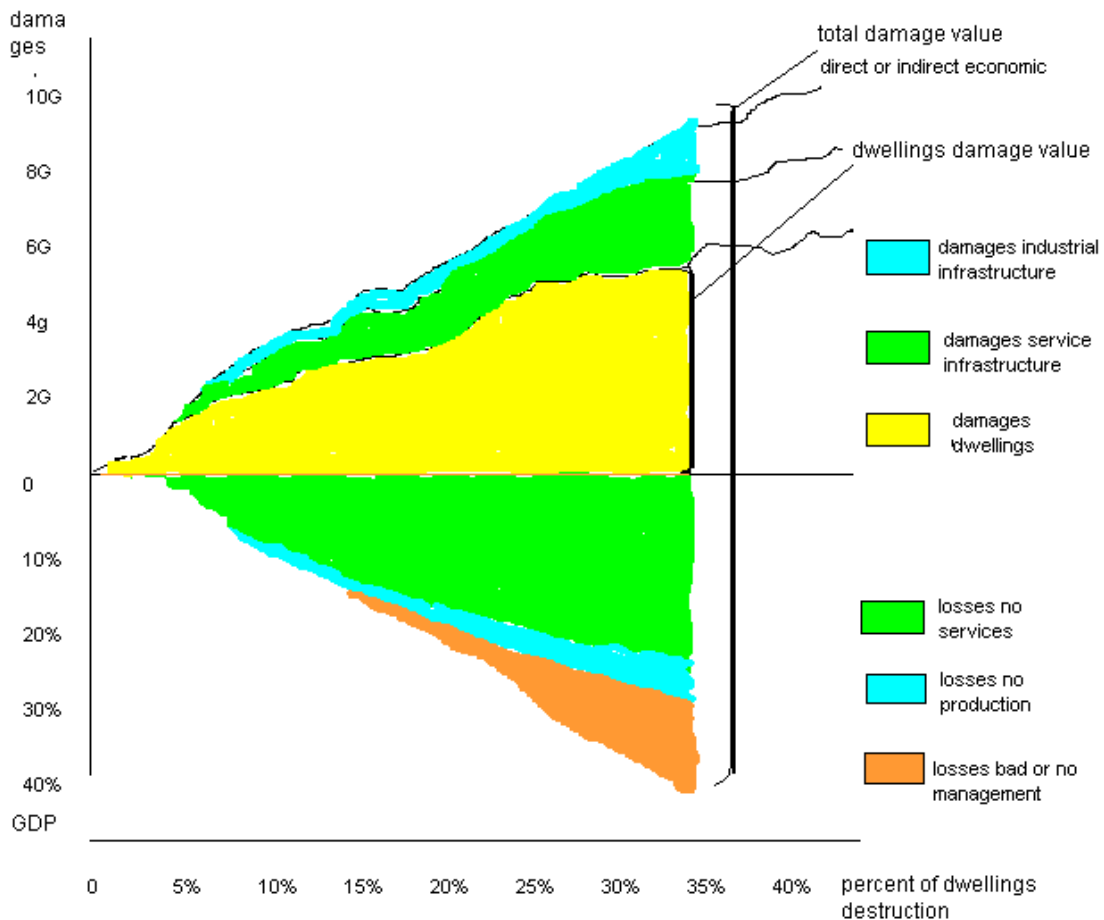


Figure 3. Non-linear, non-proportionate (as set against that of housing) increase of damage

The reasons for this are:

- In the modern societies the economy constitutes up to 70% services (commerce, public utilities, energy supply and supply of staples, education, medicine, administrative services), which are proportionate to the available housing infrastructure and which are affected/destroyed in the same degree as the housing infrastructure.
- For example, in the observation of the restoration works after the 2008 earthquake in China, destruction of a considerable part of the existing economic sector was observed on the one hand and on the other hand the replacement thereof by a newly-emerged construction sector for restoration funded by the Chinese budget.
- The industrial economic infrastructure will reduce its capacity by a given percentage of the overall damage to the housing but only within certain limits of the total disaster (for example, up to 20-30% of the housing). Individual sectors may witness less reduction when they are related to restoration of the infrastructure (construction, etc.). Above this percentage anticipatory increase of the damage of the industrial infrastructure can be expected, in parallel with the reduction of the workforce being used on the one hand (which has been redirected to salvage and restore housing), and on the other, on account of the destroyed transport infrastructure, and still other, on account of a change in the local consumption caused by the restoration boom.
- Depending on the geographical location of the industrial facilities, they too sustain damage, proportionate to that of the housing infrastructure. In the majority of industrial productions, the damage caused by a seismic disaster can cause an additional disaster: fire, pollution with chemicals, bio-agents, radioactive substances, etc., gazing, flooding, etc., which (depending on the concrete industry) can increase the damage to more than 500% from the initial value of the industrial facility.
- In large-scale disasters there inevitably occur economic losses due to a need for restructuring the economic infrastructure.

---

### **Methodology for assessment of economic damage from natural disasters**

---

This methodology constitutes an extension of the methodology for determining the total damage from a seismic disaster. In the presence of a baseline system for assessing housing damage caused by a catastrophic natural disaster (with taking into account of the "national" peculiarities), it is possible to make a multiplication scale which identifies parallel and subsequent damage on public utility and industrial infrastructure, and thus, an assessment of the relevant economic losses. A next step should analyze the cumulative effect from non-direct impacts, intangible assets, stress and psychological breakdown of big masses of the population.

A percentage of loss/destruction of public utility and industrial infrastructure commensurate to the housing infrastructure should be accepted as a basic criterion for economic loss, whereby the said infrastructures generate:

- public utility services (including transport, energy, communications, water and sewerage);
- material flows from industrial infrastructure;
- as a consequence of a reduced GDP - reduced tax revenue, restrictions/cuts in the government budget intended for restoration;
- additional reduction of the capacity for restoration and, respectively, the rates of GDP recovery.

Undoubtedly, a more precise assessment would be yielded by precise modeling of a "facility at risk" from a detailed map of the locations of public utility and industrial infrastructure compared to the map of the magnitude of the catastrophic impact via the vulnerability function, when such a map appears. At the moment information about the presence of such an integrated map is not available for any country in Europe.

As we said earlier, the public utility services are closely linked with the housing and the destruction of the former ties in very well with the damage on the latter. It is logical to recalculate the total volume of affected public utility services, via the square meters of housing area and the total lost/damaged square meters of area.

---

### Example

By way of an example we offer below the preliminary assessments about Sofia, without claiming that they will not be adjusted when additional information and up-to-date data are provided.

*Output data about Sofia, 2007*

*Gross Domestic Product 20,576,000,000 leva*

*Gross Surplus Product 16,900,000,000 leva*

*Redistributed via budget - 40% of the GDP, or 6,800 M*

*Breakdown: Services 78.5%*

*Industry 21.2%*

*Forestry and farming 0.3%*

*In services - 8.5% (increased, prior to that about 5%) construction, which will not decrease in the case of disaster.*

With a total GDP per annum (2007) of 20,576 M for the territory of Sofia City, and projected maximum damage of 32% of the housing area, the total planned economic loss from services will be 5,530 million Bulgarian leva, or about the size of the direct damage of the housing lot.

The impact on the industrial sector (extracting and processing industry) of a seismic disaster as an assessment from a correlation with damage on housing is weaker than that of the service sector. Here we first have an impact threshold (above 5-10% affected homes from a separate industrial area). Secondly, the industrial activity is concentrated in separate industrial zones, mostly outside population centers, so that heavy damage of high-rise housing would not have a significant impact on the industrial capacity. In this context it can be safely assumed that the percentage of damage of the industrial infrastructure will be no more than half of that of the housing infrastructure.

If we again take as an example Sofia City, the estimated projected economic losses would be to the tune of 411 million BG leva, or nearly a ten-times lower impact than a disaster in the service sector.

The occurrence of additional damage due to events caused by the primary disaster, such as fires, people/equipment buried under ruins, flooding, pollution, etc., has relevance only for specific industrial infrastructure. In conducting a preventive analysis, a study should be made of available major industrial or energy facilities where a fire, pollution, flooding could occur in the case of disaster.

An additional damage from a disaster in the area of Sofia could be expected in the form of above-proportionate shrinkage of wholesale trade (up to 20-30%) and mostly, above-proportionate reduction of administrative and governance services since the administration is concentrated in the area of Sofia. Then the ensuing economic effect could reach up to 20-50% of the economic losses from services, or in the case of Sofia, this would mean another 2,000 million leva in damage.

The damage from necessary restructuring of the economy following a disaster is difficult to estimate precisely enough.

---

### Conclusion

It is clear for most experts that there is little likelihood for a comprehensive methodology yielding as truthful results as possible to be devised in the next couples of years in the European Union. Examples for such failures are found in risk prevention in the United States in cases such as Hurricane Katrina, in Europe in the recent flooding in Central and Eastern Europe, the earthquake in China, etc. Thus, above all we want to promote the pragmatic step-by-step approach in the creation and application of methodologies and models for risk

management first on regional level, with a possibility for parallel verification in situ. Developing of regional focal units, engaged in a large network, covering the EU and later larger regions is a probably the best sustainable solution.

The absence of sufficient information in terms of diversity, up-to-date relevance and correctness makes it necessary to look for effective use of the available data, whereby in most cases these constitute statistical information.

Two mainstays need to be used in order to achieve good results:

- European directives and regulations;
- Strengthening the cooperation within the broad European area which includes both EU member states and neighboring countries, based on the user oriented approach which is better achieved if the regional and networking concept is accepted.

We need elaboration of national methodologies, simulation models and analytical models for damage assessment, based on local data, which are subject to constant coordination and harmonization and in direct connection with the user - national governments, civil protection units, local authorities. In this sense a main partner of the regional and national teams could be some structures of the European Commission as the Directorate General Enterprise and Industry (more particularly, the department for Competitiveness, internal market for goods and sectorial policies), Directorate General Humanitarian Aid, Directorate General Joint Research Centre (more particularly, the Institute for the Protection and Security of the Citizen).

Meanwhile, a number of countries, as well as European structures, have set up relevant websites with up-to-date information for expected or past disasters. It would be a good thing if, on the basis of peer initiatives, and in addition to the information they upload, these sites adopt a unified system for presentation of information, including:

- Automatically uploaded up-to-date information about actual events, including by regions and topics, e.g. fires, flooding, etc.; the information should be stored and accessible by time periods, e.g. weekly, monthly, yearly;
- The system should be open for inclusion of other sources of information, as well as for making comparative analyses;
- The system should also include addresses of local, national and European structures responsible for making of forecasts or for actions in the case of natural disasters.

In this presented above sense good opportunities are presented by:

- the European Directive 2007/02/EC setting the legal framework for the establishment and operation of an Infrastructure for Spatial Information in the European Community (INSPIRE), Directive 2003/98/EC on the re-use of public sector information and others;
- the creation of regional, and subsequently, European network of application servers for analyses and risk monitoring, exchanging data and information on the basis of common standards and rules for exchange and processing and which can work as a super-computer when fed with particularly large amounts of data. It is clear (figure 4), the GRID networks are convenient for solving such a problem [Kussul and Shelestov, 2008].

This can be developed jointly with several DG,s as DG ECHO, DG ENTR, DG JRC, under the Global Monitoring for Environment and Security (GMES) Program [GMES, 2010] and the future realization of Global Earth Observation System of Systems (GEOSS) [GEOSS, 2010].

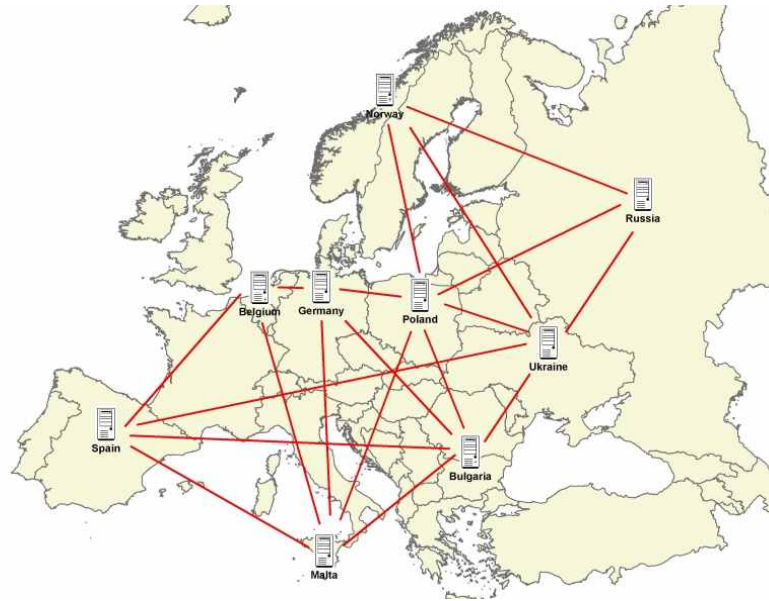


Figure 4. A structure of European network of application servers for analyses and risk monitoring

In close cooperation with the previous network, another regional network is needed to be developed - a network of regional cores (or units) for risk and security management as well as monitoring land cover changes and spatial data quality assurance, and subsequently, a European (or European-Mediterranean) network, also within the framework of the operational application of the GMES and the future GEOSS (Figure 5). The latest concept is very much open to the new tendencies and priorities in the EU 2020 strategy, as well as quality assurance (QA). It will take into account as well, the risk for the citizens, created by the ICT approaches and relevant technologies.

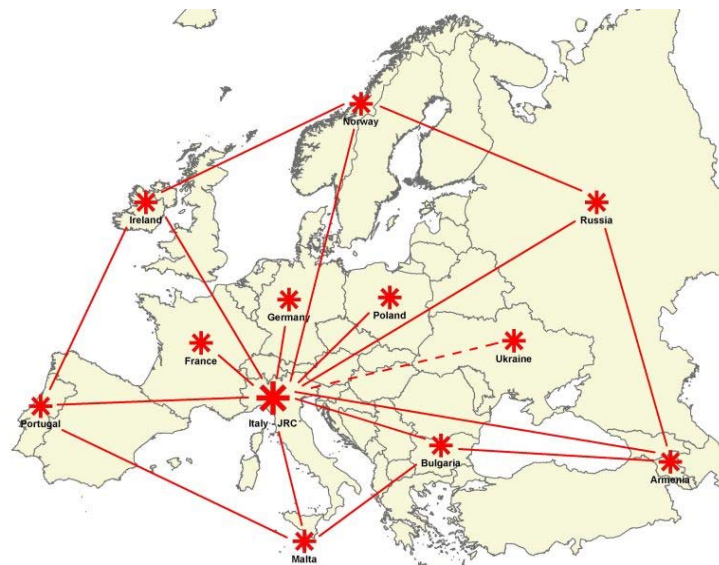


Figure 5. European network of regional units for risk and security management, land cover changes monitoring and data quality assurance

---

**Bibliography**


---

- [Declaration, 2008] "The Danube – River of the European Future", Final Declaration of the Danube Conference, Baden-Württemberg, Germany. 6th and 7th October, 2008. <http://www.european-journalists.eu/pdf/Brussels%20Danube%20Conference%20final%20declaration%20EN.pdf>
- [EU Strategy, 2010] EU Strategy for the Danube Region. [http://ec.europa.eu/regional\\_policy/cooperation/danube/index\\_en.htm](http://ec.europa.eu/regional_policy/cooperation/danube/index_en.htm)
- [GEOSS, 2010] The Global Earth Observation System of Systems <http://www.earthobservations.org/geoss.shtml>.
- [GMES, 2010] The European Earth Observation Programme (GMES). [http://ec.europa.eu/gmes/index\\_en.htm](http://ec.europa.eu/gmes/index_en.htm)
- [GMES-WS, 2010] Sofia Conclusions. GMES workshop Sofia, Bulgaria, March, 2010; <http://www.gmes-bg.org>
- [HAZUS, 2010] HAZUS. FEMA's Methodology for Estimating Potential Losses from Disasters. <http://www.fema.gov/plan/prevent/hazus/>
- [INSPIRE TR, 2007] Infrastructure for Spatial Information in Europe Reference: INSPIRE Technical Architecture Overview, 05-11-2007 Page 5 of 12
- [INSPIRE, 2007] DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007, establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union 25.4.2007 L 108/1 [http://www.epsiplus.net/content/download/3477/38314/file/I\\_10820070425en00010014.pdf](http://www.epsiplus.net/content/download/3477/38314/file/I_10820070425en00010014.pdf)
- [Kussul and Shelestov, 2008] GRID-Systems for solving the problems connected to Earth observation. Naukova Dumka, Kiev, 2008. ISBN 978-966-00-0794-9, 452p.
- [Masure and Lutoff, 2006] Masure, P., Lutoff, C. Urban System Exposure to Natural Disasters: An Integrated Approach. In: Oliveira, C.S., Roca, A., Goula, X. (ed), Assessing and Managing Earthquake Risk. Geotechnical, Geological, and Earthquake Engineering. V.2. Part III, Springer Netherlands, 2006. pp. 239-260. ISBN: 978-1-4020-3608-8.
- [Mouroux and Brun, 2006] Mouroux, P.; Brun, B. Presentation of RISK-UE Project. Bulletin of Earthquake Engineering, Volume 4, Number 4, Springer, 2006, pp. 323-339
- [Nguyen et al, 1997] Nguyen H.T., Kreinovich V., Bouchon-Meuiner B. Soft Computing Explains Heuristic Numerical Methods in Data Processing and in Logic Programming. AAAI Technical Report FS-97-04.
- [Nickolov and Polimirova, 2010] Nickolov E., Polimirova D., Critical Education Information Infrastructure's Information Security Analysis, Proceedings of the 6th Annual International Conference on Computer Science and Education in Computer Science 2010, 26-29 June 2010, Fulda/Munich, Germany, ISBN 978-954-535-573-8, pp.111-120
- [Polimirova and Nickolov, 2009] Polimirova, D., Nickolov, E., Analysis of Malicious Attacks Accomplished in Real and Virtual Environment, New Trends in Intelligent Technologies, International Book Series "Information Technologies & Computing", Number 14, 2009, ISSN 1313-0455, pp. 53-60
- [Polimirova and Nickolov, 2010] Methods and means for protection of software critical infrastructures. In: Kr. Markov and V. Velichko (ed). Intelligent Data Processing in Global Monitoring for Environment and Security, ISBN-13: 978-954-16-0045-0. ITHEA, 2010
- [Polimirova et al, 2007a] Polimirova, D., Nickolov, E., Nikolov, C., Research of the Influence of methods of compression on the co-efficient of information security of objects under attacks, International Journal "Information Theories & Applications", vol. 1 / 2007, Number 3, ISSN 1313-0455, pp.255-260.
- [Polimirova et al, 2007b] Polimirova, D., Nickolov, E., Nikolov, C., Application of methods of multi-criteria evaluation in choosing a method of compression as a means of heightening the information security of system and applied objects. International Journal "Information Theories & Applications", vol. 1 / 2007, Number 4, ISSN 1313-0455, pp. 352-358
- [Polimirova, 2008] Polimirova-Nickolova, D., Study of the Information Security of File Objects under Information Attacks with a Record of Effect of the Methods of Compression, Serdica Journal of Computing, vol. 2 / 2008, Number 3, ISSN 1312-6555, pp. 295-312
- [Wai-Wong et al, 2007] Wai-Wong K., Gedeon T., Fung C., JACIII Vol.11 No.3, 2007, pp.259-260.
- [Zielosko and Wakulicz-Deja, 2005] Zielosko B., Wakulicz-Deja A. Intelligent Data Processing in Distributed Internet Applications. In: Advances in Soft Computing, Volume 31, 2005, pp.585-591.

---

**Authors' Information**


---



**Kristian Milenov**, Assoc. Prof., PhD, Architect, Executive director of Agency for Sustainable Development and Eurointegration-ECOREGIONS, Sofia; Member of GEO HLWG;

*Major Fields of Scientific Research: Sustainable development, sustainable urban planning, energy efficiency & environmental architecture, RES, eco-auditing, GIS, remote sensing applications, risk and security management; kristian.milenov@asde-bg.org*



**Krassimir Markov** – Assoc. Prof., PhD, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. Sofia, Bulgaria. e-mail: markov@foibg.com

*Major Fields of Scientific Research: General theoretical information research, Multi-dimensional information systems*



**Venco Bojilov**, Application engineering, Business modeling: General Insurance (Bulstrad PLC BG); Risk management, INSPIRE transposition : Consultancy (ASDE BG).

*Major Fields of Scientific Research: Complex process analyze, complex functional models –ITIL (Information Technology Infrastructure Library), paradigm of risk management & mitigation, Earthquake risk management; patilan\_bojilov@hotmail.com*



**Radko Radkov**, Project manager in IT development and RS/GIS technology programs in STALKER Ltd, ASDE and ReSAC.

*Major Fields of Scientific Research: Project management in map analyses, fast track services for disaster management, creation of reliable database, control and revision by space imagery, land cover mapping, development of GIS based system for monitoring linked with control devices for dynamic data provision, satellite imagery orthorectification, DEM creation; radko.radkov@resac-bg.org*



**Pavel Milenov**, MSc in Civil Engineering. Agency for Sustainable Development and Eurointegration; Remote Sensing Application Center; National Expert at the Joint Research Centre of the European Commission in Ispra, Italy,

*Major Fields of Scientific Research: GIS and Remote Sensing, Land cover/land use mapping, image processing, land management and urban planning, infrastructure monitoring and control of the EU agriculture subsidies using satellite imagery; pavel\_milenov@yahoo.com*

## ANALOGIES BETWEEN TEXTS: MATHEMATICAL MODELS AND APPLICATIONS IN COMPUTER-ASSISTED KNOWLEDGE TESTING

Leonid Leonenko

**Abstract:** *The origin for the classification of analogy is a comparison procedure we use to make a conclusion about similarity between the source and the target. Whenever the comparison procedure is not clarified, the analogy remains ambiguous. But if this procedure is formalized the analogy may allow to formulate and check conditions of its (still partial, but often very high) confidence. Below I discuss some ideas, which grounded certain algorithms allowed to calculate different kinds of similarity between texts, in particular of natural languages. The literal similarity of two words, as well as the lexical similarity of two given texts, can be estimated using so called "Theory of Finite Sequences Similarity". The structural similarity can be measured by fixing certain name groups in the source text and check the cohesion (proximity) of names belonging to corresponding groups in the target. Special logical formalism called "The Language of Ternary Description" can provide good templates for source text structuring when comparing texts of natural languages. It was demonstrated statistically that algorithms proposed for texts' analogy estimation provide "practically trustworthy" conclusions in the knowledge testing area. I argue also that high degree of confidence for that type of analogy is connected with the background of analogy (the notion which was discussed by G. Polya) though that background need not to be evidently formalized.*

**Keywords:** *analogy, texts' similarity, knowledge testing, free answers*

**ACM Classification Keywords:** *I.2.6 Artificial Intelligence – Learning – Analogies; K.3.1 Computers and Education - Computer Uses in Education*

---

### Introduction

---

I assume the following general definition of analogical inference [Leonenko, 2008a]:

$$L: \approx_{C_p}(a,b) \in \Omega, P(a) \vdash Q(b)$$

where  $a$  is the source of analogy;  $b$  is the target;  $L$  is a language using for representing  $a$  and  $b$ ;  $C_p$  is a procedure (algorithm) of comparison that estimates the degree of their similarity;  $\Omega$  is a set of similarity values;  $P$  and  $Q$  are  $L$ -descriptions which characterize  $a$  and  $b$ ;  $\approx_{C_p}(a,b) \in \Omega$  means that similarity degree of  $a$  and  $b$  belongs to  $\Omega$ . The procedure  $C_p$  serves as a basis for analogy types' classification.

Let the source  $a$  be a correct answer for test question, and let the target  $b$  be a variant of answer coming from the student.  $(P,Q)$ -pair corresponds to the information transferred from the source to the target, here it will be the correctness or admissibility of the answer (which is reflected by the "student's mark"). The restriction comes from computer is that both source and target must be represented as symbol sequences, or "texts" (if the answer is entered in non-textual form, the computer program will still deal with some symbolic representation of it). Analogical inferences take place in all cases when the student is allowed to use free form of the answer. Thus multiple choice is not in the case. But for "open" answers different comparison procedures  $C_p$  can be accepted for different presupposed structures of source and target objects.

For instance, if the answer is expected to be a set of numbers and the student's mark is calculated as a measure of intersection of this set and a "correct" one, then we obtain, factually, the traditional type of analogy – Aristotelian "paradeigma" [Uyemov, 1970]. But we meet with more complicated analogy types if take into consideration: 1) the differences in the used notions of numbers' identity (do we treat as identical real and integer numbers of the same value; or real numbers that are equal up to some accuracy; etc.); 2) the order in the two compared sets of numbers; 3) the condition of compactness for some subsets in the compared sets; and so on.

The other analogies appear when a computer program needs to recognize a mathematical expression in student's answer. Here the type of analogy depends on: 1) the kind of a given expression (e.g.: function, equation, identity, etc.); 2) the mode of expression's representation (e.g.: in some algorithmic language, or with



given graphical primitives); 3) the equivalence class which is presupposed for the given kind of expressions – it determines cases when two different expressions must be considered as equivalent (here “equivalence” is the mode of interpreting “similarity”). For example, if the student’s answer is a function represented as  $b(x)$ , we may assume that it can be accepted as correct if it coincides with the sample function  $a(x)$  for one (or two, or larger, but finite number) argument value  $x_0$ :  $b(x_0) = a(x_0)$ .

### Analogies between “texts” composed of “words” organized into “groups”

The most interesting are analogies using to recognize answers represented in natural languages (and also mixed answers, when student is allowed to use natural language together with mathematical expressions or other “artificial words” to express the answer). Here the assuming concept of “an object of  $L$  language” is “a text consisted of words”. Let us assume the following terminology. Two words are called “similar” if there exists such common subsequence of their symbols which has “appropriate” length. Two texts are similar when contain many similar words. In spite of the simplicity of these initial concepts the mathematical model for that kind of analogy had required proving some nontrivial theorems. The most interesting is the fact that it provides very good results in real computer assisted testing processes. Answers that require natural language can be estimated by computer with necessary accuracy, and nonessential errors in answer can be disregarded.

Let me present one example. Imagine that we need to compare the following two sentences:

***Newton and Leibniz were inventors of the mathematical analysis*** (S)

***The math. analysys was Newton’s and Leibnitz’s envention*** (T1)

Let **S** be an “etalon” or “standard” answer for the question

*What was the most important achievement in mathematics at the turn of the 17-18 c., and who was its author ?* (Q)

**Q** is posed to the student in the computer assisted knowledge testing session; and the phrase **T1** is a factual answer coming from the student. **T1** contains several mistakes and changes in words (underlined). The order of words in **T1** differs from that in **S**. But it is clear that **T1** is a correct answer to the question **Q**.

We see that there exists certain similarities between some words of **S** and **T1**, e.g.:

{*inventors* , *envention*} or {*analysis* , *analysys*},

– two words in each pair are “analogical”.

One can say that “analogical” is not appropriate word here. But let me consider any word in the phrase as “an object”=“a complex of attributes”, where each “attribute” is an *occurrence* of a symbol in that word (i.e. a pair <symbol, position>). The position is treated as not a numerical index, but rather a relation between the set of previous and the set of next occurrences. Two words are considered as having “the same attribute” if for a given occurrence of their common symbol two sequences of letters that stand before/after that symbol are identical in each word. Furthermore, two words are considered as having “similar attributes” if the mentioned sequences “slightly differ” one from another. Finally, two words are considered to be the more similar the more “similar attributes” they have. Thus we come to something very closed to the traditional analogy of Aristotle (paradeigma), though our notion of “attribute” includes references to other “attributes”, and so factually we have a mix of paradeigma and proportional analogy.

Specializing the above ideas, there can be introduced measure function  $F(a,b)$  that estimates analogy between two words  $a$  and  $b$  [Leonenko & Poddubny, 1996]. It can be shown that this measure produces the correspondent notion of the “distance” between words:

**Theorem 1.** The function  $1-F(a,b)$  forms a metric space in the set of words.

We can make further assumptions to precise the notion of words' similarity in the considered case. E.g., we can postulate that some letters in a word can be more "essential" than others, and assign certain "weight" to each symbol in a word:

i n v e n t o r s  
1 1 1 1 1 1 0 0 0

There can be proposed measure function  $G(a,b,W)$  to calculate "the degree of analogy" treating as weighted similarity of words [Leonenko, 2002]. Applying it, we will obtain, e.g., the value **0.833** for the pair of words

**{inventors, envention}**

The third parameter  $W$  in  $G(a,b,W)$  reflects the "expecting potential weight" of those letters of the target  $b$  which are not present in the source  $a$  and therefore can't be weighted directly. Thus  $G$ , unlike  $F$ , is not a symmetric function with respect to  $a$  and  $b$ . Nevertheless,  $G$  has the following "natural" properties:

**Theorem 2.** (1) if  $W_1 \leq W_2$  then  $G(a,b,W_2) \leq G(a,b,W_1)$ ;

(2)  $0 \leq G(a,b,W) \leq 1$  for every  $a,b,W$ ;

(3) if  $W > 0$  and  $G(a,b,W) = 1$  then  $b$  coincides with  $a$  up to some letters that have zero weights;

Passing from words to phrases, we see that there exists a similarity between "source" phrase **S** and "target" phrase **T1**: they are "analogical" in the sense that have many "analogical" words; – just like two words that have many common symbols, but with the addition that the *order* of words in phrases is *less rigid* than the order of letters in words. There was introduced the measure function  $G_U(a,b,W)$  that calculates weighted similarity *fully ignoring* the order. So if we assign, e.g., the following weights to words of **S**:

**Newton and Leibniz were inventors of the mathematical analysis**

25    0    25    0    10    0    0    10    30

then  $G_U(\mathbf{S}, \mathbf{T1}, W)$  gets the value **1.00**, thus with respect to  $G_U$  **S** and **T1** appear as "fully analogical" phrases. There exist some "natural" properties and relations between  $G(a,b,W)$  and  $G_U(a,b,W)$  measures:

**Theorem 3.** (1) if  $W > 0$  and  $G_U(a,b,W) = 1$  then  $b$  is a permutation of those sub-objects of  $a$  that have non-zero weights;

(2)  $G(a,b,W) \leq G_U(a,b,W)$  for every  $a, b$  and  $W$ .

The next step towards more accurate estimation of analogy between phrases can take into account certain structures in phrase. Consider, e.g., the following combination of words:

**The mathematics of Newton invents Leibniz's analysis** (T2)

Though **T2** includes analogs of *all* non-zero weighted words of the "source" phrase **S**, **T2** does not look like a correct answer to the question **Q**. So we need to account that shuffling of words in **T2** affects in decreasing its similarity to **S**. There can be proposed the measure for "the cohesion" of given group of words in the phrase [Leonenko, 2007]. It is obvious, that in both **S** and **T1** each of the two following groups – **{Newton, Leibniz}** and **{mathematical, analysis}** – are cohesive, while in **T2** they are not.

Let  $a$  and  $b$  be some words belonging to the same group  $g$ . In order to formalize the notion of "cohesive group" we may start with the definition of the "measure of disconnection" of words  $a$  and  $b$  in a given text  $T$ . It is defined as a number  $\rho(a,b)$  of those words, which occur in  $T$  between  $a$  и  $b$  and not belong to  $g$ . It is easy to see that  $\rho(a,b)$  is a distance function on the subset of those elements of  $g$  which occur in  $T$ . Let  $T$  contain exactly  $v$  occurrences of group  $g$ 's elements, and let  $v \geq 1$ . Assign the indexes  $1, \dots, v$  to these occurrences preserving their order in  $T$ ; and designate the obtained set  $\{c_1, \dots, c_v\}$  as  $g[T]$ . In the case of  $v > 1$  I will call "the disconnection matrix" of the set  $g[T]$  the  $v \times v$ -dimensional matrix  $M$ , whose elements  $M_{q,r}$  are equal to  $\rho(c_q, c_r)$  in the sequence  $T$ .

"The average disconnection"  $\mu(g[T])$  of the set  $g[T]$  is defined as zero when  $v=1$ , and otherwise ( $v > 1$ ) as:

$$\mu(\mathbf{g}[\mathbf{T}]) = \frac{\sum_{q=1}^v \sum_{r=1}^v M_{q,r}}{v(v-1)},$$

where  $\mathbf{M}$  is the disconnection matrix of  $\mathbf{g}[\mathbf{T}]$ .

“The spectrum of the disconnection” of the set  $\mathbf{g}[\mathbf{T}]$  is defined as the sequence of natural numbers  $L(\mathbf{g}[\mathbf{T}]) = \langle L_1, \dots, L_{v-1} \rangle$ , where  $L_k = \rho(c_k, c_{k+1})$ ,  $k \in [1, v-1]$ . The following theorem shows that it is easy to calculate  $\mu(\mathbf{g}[\mathbf{T}])$  using the spectrum of the disconnection:

**Theorem 4.** 
$$\mu(\mathbf{g}[\mathbf{T}]) = \frac{2}{v(v-1)} \sum_{j=1}^{v-1} j(v-j)L_j.$$

Finally, let me introduce the notion of so called “ $\varphi$ -cohesion”. Let  $\varphi$  be an arbitrary function with the following properties:

- 0)  $\varphi(x)$  is determined for every  $x \geq 0$ ;
- 1)  $\varphi(0) = 1$ ;
- 2)  $\varphi(x)$  is non-increasing for  $x \rightarrow \infty$ .

Then  $\varphi$ -cohesion of group  $\mathbf{g}$  elements in the text  $\mathbf{T}$  is defined as  $\varphi(\mu(\mathbf{g}[\mathbf{T}]))$ .

The  $\varphi$ -cohesion can be used as the specific measure of similarity for those subsequences (groups) of words in the sentence, which are somehow “clogged” with alien, extraneous words. This type of similarity may be resembled to the notion of “recognizing of the sub-phrase”. The condition 1), which must be satisfied by  $\varphi$ , means that if there are no extraneous words between words from  $\mathbf{g}$  in  $\mathbf{T}$ , then the corresponding sub-phrase can be “fully recognized”. Otherwise the degree of its recognition can be (not obligatory) less than 1.

The concrete form of the function  $\varphi$  can be chosen under the accepted hypothesis of how the ability to recognize the given sub-phrase decreases when this sub-phrase becomes “clogged”. If one accepts that the only one extraneous word strongly affects the recognition, then it may be suitable to choose  $\varphi = e^{-x}$ . But if it is accepted that a sub-phrase can be “half-recognized” independently of the number of words that “clog” it, one can chose  $\varphi = (1+x)/(1+2x)$ , which has the limit 1/2 when  $x \rightarrow \infty$ .

Let us look how the above measure of  $\varphi$ -cohesion works being applied to the English sentences considered above. We must first divide the standard sentence  $\mathbf{S}$  into the nonintersecting sets of words. Let

$\mathbf{g}_1$  be the set of words, which are analogous (i.e. similar in respect of common letters, their weights and order) to one of the two words: **Newton** or **Leibniz** on the appropriate similarity levels;

$\mathbf{g}_2$  be the set of words, which are analogous to one of the words: **inventors**, **founders**, **devisers**, and so on;

Now when we will compare the standard phrase  $\mathbf{S}$  to an arbitrary phrase  $\mathbf{T}$ , we will form sets  $\mathbf{g}_i[\mathbf{T}]$  as  $\mathbf{g}_i[\mathbf{T}] = \mathbf{T} \cap \mathbf{g}_i$ ,  $i \in [1, 3]$ .

If the order of words in phrases  $\mathbf{S}$  and  $\mathbf{T}$  is *not* taken into consideration, the similarity between these two phrases is proportional to the total weight of their common words. Let us now assume that the total weight of the group  $\mathbf{g}_i[\mathbf{T}]$  *must be decreased by multiplying it by the  $\varphi$ -cohesion of  $\mathbf{g}_i[\mathbf{T}]$  in  $\mathbf{T}$ .*

Then, if we will take  $\varphi(x) = 1/(1+x)$  to calculate the  $\varphi$ -cohesion, we will obtain the following results:

	The value of the weighted similarity to $\mathbf{S}$ when $\varphi$ -cohesion is ignored	The value of the weighted similarity to $\mathbf{S}$ when $\varphi$ -cohesion is taken into account
<b>T1</b>	1.00	1.00
<b>T2</b>	1.00	0.45

Thus using the measure of cohesion *together* with the measure of weighted similarity, we decrease the value for the analogy between **S** and **T2**, preserving it for **S** and **T1**.

The last problem I will discuss in this section is the problem of fixing the nonintersecting groups of words in a given “standard” answer. If we do not leave this problem to the intuition of a person who prepares the testing question, we may turn either to the linguistics of a given natural language or to some logical analysis of a phrase. The last way presupposes some formal logical apparatus. It is well-known that different logical calculi have different abilities in reflecting characteristics of natural phrases. Let me notice the logical calculus called “The Language of Ternary Description” (LTD), elaborated by A. Uyemov [Uyemov, 1995]. The “source” sentence **S** considered above has the following LTD-model:

$$i(*\{\{N, L\} \bullet [(a)m]\}),$$

where *i* denotes “invention”; **N** – “Newton”; **L** – “Leibniz”; **a** – “analysis”; **m** – “mathematical”.

There was proposed the hypothesis that cohesive groups in natural language’s sentence correspond to the nonintersecting 1st level sub-formulas of the LTD-model of that sentence [Leonenko, 2008b]. For the above sentence **S** these will be sub-formulas *i*, **{N, L}** and **[(a)m]**. Note that parentheses and square brackets in LTD denote special operations of predication and “reification”, that’s why *i* is a separate sub-formula in  $i(*\{\{N, L\} \bullet [(a)m]\})$ . (The function of arrangement is played by the braces.)

When the hypothesis stated above is used in the processes of calculating the  $\varphi$ -cohesion, it is assumed, roughly speaking, that words of the given LTD-*vicinity* must be closed enough one to another in the natural sentence. This assumption was confirmed by testing different natural sentences (mainly from Ukrainian and Russian languages). In particular, we can see that it works for the above English sentences **S**, **T1** and **T2**.

It can be showed that algorithms which estimate different types of texts’ similarities using the above measure functions have polynomial computational complexity.

The described methods of evaluating analogies between sentences were checked on the large amount of texts in computer assisted knowledge testing sessions. They showed very good practical results. In fact, it was statistically proved that these measures for analogies give practically same conclusions about student’s attainments to those obtained by manually checking written tests allowing “open” answers of medium length [Baranov, 2004].

---

### Overcoming the “gaps of inference” to get trustworthy conclusions of analogy

---

Though statistical confirmation is important, there still remains the following problem. Let a computer had estimated *concrete* student’s answer **T** to some *concrete* question as highly similar to an “standard” answer. What must be the degree of our confidence in the correctness of the answer **T**?

This problem is connected with the general problem of the validity of analogical inferences. I will discuss some aspects of it in this last section.

At first, I want to draw attention to the following fact. The notion of “an attribute” in examples of analogies between texts given above had become very specific. As far as I acquainted with different theoretical and practical approaches to analogy (e.g., in probabilistic logic, structure mapping, expert systems, etc. – see [Niiniluoto, 1988], [Gentner & Kurtz, 2006], [Kokinov & French, 2003], [Spanoudakis & Constantopoulos, 1996]), the deep specifications of the assumed notions of “an object”, “an attribute of object”, “a relation between objects” and “a similarity between objects” appear as necessary conditions in all the cases when the conclusion of analogy has high degree of certitude. In most of these cases one can make fully transparent his claim that two objects are “similar”, and usually can present the algorithm to compare the source and the target and evaluate their similarity.

The so-called “validity principles of analogy” [Uyemov, 1971] in situations when the comparison procedure can be make precise transform to special rules (sub-algorithms) of comparison. Thus, the traditional rule of paradeigma “take the most important attributes” transformed in examples in previous section into “select analogs with the high total weight”.

When comparison procedure can't be described precisely the validity principles become fuzzy, like "take as many properties as possible" in classic paradigm. The last principle can even be unacceptable in some precise types of paradigm (in previous examples another principle works: "do not take properties of zero weight"). If the rules for comparing objects are fuzzy, we can't be sure that two given objects are "really" similar. This is our main reason to say that the conclusion of analogy is "only questionable".

So I insist that in general case the existence of an algorithm that can *evaluate* the *degree* of similarity between the source and the target is the *necessary condition* for analogy to produce confident results.

But what about the *sufficient* conditions? When the comparison procedure is precise and gives high degree of similarity, we often rightfully trust the conclusion of analogy. Why can we do this?

Sometimes it happens because the conclusion is deductively entailed from the similarity of the source and the target (like in strict isomorphism). But in most cases there exists a "gap of deduction" in making analogical conclusions. I now discuss one case when we can ignore this gap.

Let me return to the area of computer assisted knowledge testing. Consider the following kind of analogy used to evaluate the answer that is a mathematical function, e.g.  $f(x)=0.35\sin 2x$ . It is obvious that if the student will enter any formula equivalent to  $f(x)$ , like  $g(x)=0.7\sin x \cos x$ , his answer will be correct. The usual way for computer to check such answers is to compare the values of  $f(x)$  and  $g(x)$  for one (or two, or larger, but *finite* number) argument value  $x_0$ :  $f(x_0) = g(x_0)$ .

If we fix the appropriate value for  $x_0$  (like  $x_0=0.314$  for the above example), we can be *absolutely sure* that the coincidence of the student's and the "standard" numerical results guarantee the correctness of his answer. But this kind of logical inference can be treated as a paradigm with the *single* attribute of object (or, if you wish, as an enumerative induction with the single trial). Traditional "validity principles" for paradigm (or induction) do not recommend accepting such conclusions. In fact, there exists an infinity of functions that coincides with  $f(x_0)$  on the argument  $x_0=0.314$ .

I think that high degree of confidence we tend to assume for that type of analogy is connected with the *background* of analogy (see [Polya, 1954]). The student (even if he knows or guesses the method of comparing his and standard answer) needs to know the concrete value of  $x_0$  to "deceive" the computer. This is almost impossible in the considered situation (i.e. on the given background of analogy).

It can be said that the precision of the comparison procedure of the source and the target of analogy must be brought up to the extent when the background of analogy will give us the possibility to overcome the "gap of deduction". E.g., in the case of comparing textual student's answers we come to the "1-st approximation" of such a background when guarantee that "most of *essential* words of standard answer *somehow* occur in the student's one". The next step may be done when we can check if "most groups of names that *are cohesive* in the standard answer *are cohesive* in the student's one", and so on.

In other words, the sufficient conditions for the validity of analogy are related to those – maybe, extra-logical – circumstances (background), which allow to assume that when taking the target with high similarity to the source we can (under this background) perform the transfer of information.

The practical examples discussed above show that in some cases we may fix the background of analogy to ensure trustworthy conclusions, though this background includes extra-logical assumptions.

---

## Conclusion

There exist different aspects in which two given texts of given natural or artificial language can be treated as being "analogical". If we consider "text" as a "sequence of phrases", and "phrase" as a "sequence of words", then the Theory of Finite Sequences Similarity gives us tools to estimate degrees of literal similarity between words and lexical similarity between phrases or texts. We can also use logical calculi appropriate to fix "cohesive" structures in natural languages' texts, and then apply mathematical measures to check the cohesion of a given name group in text being the target of analogy. It gives the possibility to estimate the similarity of texts' structures. Taking together, these methods can make a contribution, in particular, to the problem of "free" answers in

computer-assisted knowledge testing. Namely, the algorithms discussed above are able to disregard nonessential lexical and structural changes in a given “free” answer. As a result, estimations of “free” answers produced by computer do not differ from those of humans. It was confirmed by statistical experiments, but it also can be justified by taking into account the specific background of analogies in automatic knowledge testing area.

---

## Bibliography

- [Baranov, 2004] V. Yu. Baranov. Computer-assisted testing in informatics: the results of pedagogical experiment in Odessa National Academy of Telecommunications // Theory and methods of education in mathematics, physics, informatics. Issue 4, Vol.3. – Krivyi Rig, 2004, p.6-12. (In Ukrainian).
- [Gentner & Kurtz, 2006] D. Gentner, K.J. Kurtz. Relations, Objects, and the Composition of Analogies // Cognitive Science, Vol.30, 2006, p.609–642.
- [Kokinov & French, 2003] B. Kokinov, R.M. French. Computational Models of Analogy-making // Encyclopedia of Cognitive Science. Vol. 1. London: Nature Publ. Group, 2003, p.113-118.
- [Leonenko, 2002] L. Leonenko. Analogical inferences in computer assisted knowledge testing systems // 6th Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2002), Orlando, Florida, USA, 2002. Proceedings, Vol. XVIII, p.371-376.
- [Leonenko, 2007] L. Leonenko. Algorithms for estimation analogies between texts and its applications in computer assisted knowledge testing // VII Intern. Conf. “Intellectual Analysis of Information”. Transactions. – Kyiv: Prosvita, 2007, p. 210-220. (In Russian)
- [Leonenko, 2008a] L. Leonenko. Analogy and Abduction // Filosofska Dumka, 2008, No. 3, p.14-30. (In Ukrainian)
- [Leonenko, 2008b] L. Leonenko. The Language of Ternary Description in Text’s Cohesion Estimations // VIII Intern. Conf. “Intellectual Analysis of Information”. Transactions. – Kyiv: Prosvita, 2008, p. 286-295. (In Russian)
- [Leonenko & Poddubny, 1996] L. Leonenko. H. Poddubny. The Theory of Finite Sequences Similarity and its Applications to Image Recognition // Automatics and Telemechanics, 1996, No.8, p.119-131. (In Russian)
- [Niiniluoto, 1988] I. Niiniluoto. Analogy and Similarity in Scientific Reasoning // Analogical Reasoning. – Synthese library, Vol.197, Kluwer Acad. Publishers, 1988, p.271-298.
- [Polya, 1954] Polya G. Mathematics and Plausible Reasoning.– Princeton Univ.Press, NJ, 1954.
- [Spanoudakis & Constantopoulos, 1996] G. Spanoudakis, P. Constantopoulos. Elaborating Analogies from Conceptual Models // International Journal of Intelligent Systems, 1996, Vol.11, No.11, p.917-974.
- [Uyemov, 1970] A. Uyemov. Analogy in the Practice of Scientific Researches. – “Nauka” Press, Moscow, 1970. (in Russian)
- [Uyemov, 1971] A. Uyemov. The logical foundations of modeling method. – Moscow: “Mysl” Press, 1971. (In Russian)
- [Uyemov, 1995] A. Uyemov. The Language of Ternary Description as a Deviant Logic // Boletim da Sociedade Paranaense de Matematica. – 1995, Vol.15, No.1-2, p.25-35; 1997, Vol.17, No.1-2, p.71-81; 1998, Vol.18, No.1-2, p.173-190.

---

## Authors' Information



**Leonid Leonenko** – Odessa National Academy of Telecommunications, Associate Professor, Kinna str., 22, apt.6, Odessa, 65029, Ukraine; e-mail: [Leonid.Leonenko@gmail.com](mailto:Leonid.Leonenko@gmail.com)  
 Major Fields of Scientific Research: Non-classical logic, General system’s theories, Computer-assisted knowledge testing

## ON SEMANTICS AND SYNTAX OF THE BSDT PRIMARY LANGUAGE

Petro Gopych

**Abstract:** Within the binary signal detection theory (BSDT) the semantics and syntax of a primary language (PL, a mathematical framework for internal brain computations) have been proposed and described in a semi-formal form. On the basis of BSDT infinity hypothesis (the infinity of common in the past prehistory of universe, life, and mind), basic BSDT PL notions have been defined. Among them the names of real-world things, their meanings/contexts (they are infinite and common in the past binary strings), meaning complexity, truth, and understanding the truth. Given their infinite contexts BSDT PL names are finite-in-length binary strings that may simultaneously be interpreted as Gödel numbers or uncomputable halting probabilities (fractions of Chaitin's  $\Omega$ ) for binary string algorithms running on particular self-delimiting computers. BSDT PL meaning complexity is compared with Shannon entropy/information, Kolmogorov/algorithmic complexity, Gell-Mann and Lloyd's effective complexity and total information; BSDT PL truth is compared with Tarskian truth. High biological plausibility of the BSDT PL, its potential for disigning the languages with capacities at the level of human natural languages, applications to practical semantic computations and testable empirical predictions are discussed. Because of its infinity hypothesis, BSDT PL is beyond the scope of traditional axiomatic approach to logic and mathematics.

**Keywords:** meaning, complexity, truth, symbolic communications, semantic computations.

**ACM Classification Keywords:** C.3 Special-purpose and Application-based Systems; E.4 Coding and Information Theory; F.1.3 Complexity Measures and Classes; H.1.1 Systems and Information Theory; I.2.0 General, I.2.4 Knowledge Representation Formalisms and Methods; J.4 Social and Behavioral Sciences

### 1. Introduction

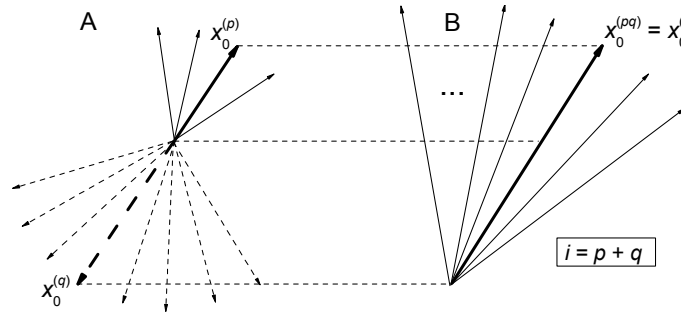
Since Frege [Frege, 1892] the problem of meaning or reference has been regarded as the central problem in investigations of language. At the same time, it is supposed 'we will not get an adequate theory of linguistic reference until we can show such a theory is part of the general theory ... of how the mind is related to objects in the world in general' [Searle, 1981; p. xi]. Recent binary signal detection theory (BSDT [Gopych, 2008a]) and its atom of consciousness model (AOCM [Gopych, 2009b]) provide the background for the theory required. On this basis a possibility arises to find a new solution to the problem of meaning/interpretation – the main focus of semantics. In this paper such a solution is proposed using the BSDT version of a low-level or 'primary language truly used by the central nervous system' and structurally 'essentially different from those languages to which our common experience refers' [von Neumann, 1958; p. 92].

BSDT primary language (PL) gives a common mathematical framework for the description of brain spiking activity involved in signal processing, memory storage/retrieval, decision-making [Gopych, 2008b], and consciousness [Gopych, 2009b]. It is supposed it has to be sufficient to maintain the basic/principal/involuntary behavioral tasks that do not require for their support any additionally elaborated symbolic interanimal communication systems. It may also serve as a ground for a 'bottom-up' description and design of more complicate symbolic systems, including human natural/public/ordinal word languages, used by animals/humans for their social interactions.

### 2. BSDT PL Vocabulary, Infinity Hypothesis

The BSDT operates with  $i$ -dimensional spinlike (with components  $\pm 1$ ) binary vectors  $x^{(i)}$  constituting a binary vector space,  $S_{x_i}$  [Gopych, 2009a]. They may also be considered as  $i$ -bit strings or elements/points of the set,  $S_{x_i}$ , whose cardinality (the number of elements) is  $|S_{x_i}| = 2^i$ . If there are string variables  $x^{(p)}$ , with values ranged in  $S_{x_p}$ , and  $x^{(q)}$ , with values ranged in  $S_{x_q}$ , then  $x^{(i)} = x^{(p,q)} = x^{(p)}x^{(q)}$ , with values ranged in  $S_{x_i} = S_{x_pq}$ , can be treated as 'compound' variables;  $|S_{x_i}| = 2^i$ ,  $i = p + q$ ; if  $p \leq q$ ,  $S_{x_p} \subseteq S_{x_q}$ . The space  $S_{x_i}$  may also be interpreted as either the  $S_{x_p}$  whose vectors are colored in  $2^q$  colors or the  $S_{x_q}$  whose vectors are colored in  $2^p$  colors. If so,  $p$  and  $q$  are the measures of discrete ('colored') nonlocalities ('rainbows') of points/vectors in spaces  $S_{x_q}$  and  $S_{x_p}$ , respectively.

Two-color nonlocality for coding/decoding of noised binary signals was earlier analyzed in [Gopych, 2009a]. By induction it can be demonstrated that vectors related to any number of such spaces can similar be combined (see Fig. 1). For compound strings, communicative and associative laws are valid, namely  $x^{(pq)} = x^{(qp)}$ ,  $x^{(pq)} = x^{(p)}x^{(q)}$  =  $x^{(p)}x^{(q)}$ , etc. BSDT vectors/strings just introduced can be thought of as *words* or names/terms of BSDT *primary language* (PL). They constitute its *vocabulary* written using the PL *alphabet*: '+1', '-1', usual signs of arithmetics, and some auxiliary simbols.



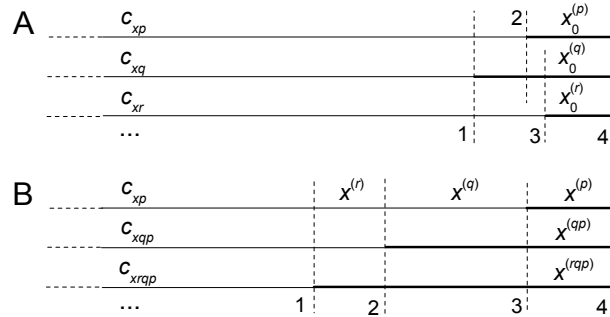
**Figure 1.** The scheme of designing a compound space,  $S_{xpq} = S_{xi}$ . **A**, Spaces  $S_{xp}$  and  $S_{xq}$ , the constituents of  $S_{xi}$ ; string variables  $x^{(p)}$  ranged in  $S_{xp}$  and  $x^{(q)}$  ranged in  $S_{xq}$  are respectively shown as solid and dashed arrows ( $p = 2$ ,  $q = 3$ ); particular values of  $x^{(p)}$ ,  $x_0^{(p)}$ , and  $x^{(q)}$ ,  $x_0^{(q)}$ , are highlighted in bold; the set of dashed arrows displays the 'rainbow' of  $|S_{xq}| = 2^q = 8$  'colors' that may color each of the solid arrows ( $2^q$ -nonlocality of vectors in  $S_{xp}$ ), or analogously for dashed arrows ( $2^p$ -nonlocality of vectors in  $S_{xq}$ ). **B**, Compound space  $S_{xi}$  (it is uncolored,  $i = 5$ ,  $|S_{xi}| = 2^i = 32$ ); variables  $x^{(i)}$  ranged in  $S_{xi}$  are shown as in **A** (a part of them is only presented); particular  $x^{(i)} = x_0^{(i)}$  is highlighted in bold. Horizontal dashed lines point to relations between the lengths of  $x_0^{(p)}$ ,  $x_0^{(q)}$ , and  $x_0^{(i)}$ .

Let  $c_{x0}$  be a one-side infinite (infinite 'in the past', at the beginning) spinlike binary string of the length  $l(c_{x0})$ . If  $c_{x0}$  is arbitrary disjointed into two parts, we get  $c_{x0} = c_{xp}x_0^{(p)}$  where  $c_{xp}$  is again infinite in the past string while  $x_0^{(p)}$  is a finite string of the length  $l(x_0^{(p)}) = p$ . Then the length of  $c_{xp}$  equals  $l(c_{xp}) = l(c_{x0}) - p$ . We say that compound string  $c_{x0} = c_{xp}x_0^{(p)}$  is a value of (or bound condition for) the string form  $c_{xp}x^{(p)}$  with  $c_{xp}$  as a string constant and  $x^{(p)}$  as a string variable ranged in  $S_{xp}$ . This form provides  $|S_{xp}|$  compound strings of the length  $l(c_{x0})$  that share their infinite initial part,  $c_{xp}$ , prefixed by different values of  $x^{(p)}$ . If  $c_{x0}$  is disjointed into two parts in another way, e.g.  $c_{x0} = c_{xq}x_0^{(q)}$ , then we similarly obtain the form  $c_{xq}x^{(q)}$  where  $c_{xq}$  is an infinite string constant of the length  $l(c_{xq}) = l(c_{x0}) - q$  and  $x^{(q)}$  is a string variable of the length  $l(x^{(q)}) = q$  ranged in  $S_{xq}$ . The number of compound strings of the length  $l(c_{x0})$  shared infinite initial part  $c_{xq}$  that is prefixed by different values of  $x^{(q)}$  is  $|S_{xq}|$ . Since  $c_{xp}$ ,  $c_{xq}$ ,  $c_{x0}$  share their infinite initial parts and  $c_{xp}$  and  $c_{xq}$  are prefixed given the length of  $c_{x0}$  by prefixes of different non-zero lengths,  $p \neq q$ , either  $c_{xp}$  is longer than  $c_{xq}$  by  $\text{abs}(p - q)$  bits or vice versa (Fig. 2A). Hence, for the above reasons and under above conditions, in spite of one-side infinity of  $c_{x0}$ ,  $c_{xp}$ , and  $c_{xq}$ , their lengths can explicitly be compared. That is the BSDT *infinity hypothesis* which is rather well empirically supported and reflects the unity of universe, life, and mind [Gopych, 2009a]. (Note, the number of ways the  $c_{x0}$  can be disjointed into two parts is the length of  $c_{x0}$  in bits,  $l(c_{x0})$ , or the first Cantor's infinite cardinal number.)

The above consideration can be generalized by assuming the existence of a set,  $S_{cx0}$ , of all possible infinite and common in the past compound strings  $c_{xi}x_j^{(i)}$  of the length  $l(c_{x0})$ . Here,  $c_{xi}$  is infinite in the past string of the length  $l(c_{xi})$ ,  $x_j^{(i)}$  is its  $j$ th prefix of the length  $i$ , and the (bounded) condition  $l(c_{x0}) = l(c_{xi}) + i$  holds. Index  $i = 0, 1, 2, \dots$  specifies the ways of dividing the  $c_{x0}$  or generating the  $c_{xi}$  (if  $c_{x0}$  remains intact,  $i = 0$ ,  $c_{xi} = c_{x0}$ ); index  $j = 1, 2, 3, \dots, 2^i$  specifies the  $j$ th value of string variable  $x^{(i)}$ ,  $x_j^{(i)}$ . Mentioned above strings  $x_0^{(p)}$ ,  $x_0^{(q)}$ ,  $x_0^{(i)}$  are examples of values  $x_j^{(p)}$ ,  $x_j^{(q)}$ ,  $x_j^{(i)}$ . Every  $c_{xi}x_j^{(i)}$  related to the  $S_{cx0}$ ,  $c_{xi}x_j^{(i)} \in S_{cx0}$ , is uniquely labeled by decimal indices  $i$  and  $j$  or by the string  $x_j^{(i)}$  (it encodes  $i$  and  $j$  in binary notations as its length and its content, or particular arrangement of its positive and negative components; prefixes/names  $x_j^{(i)}$  provide actually *Gödel numbering* of strings  $c_{xi}x_j^{(i)}$ , the members of  $S_{cx0}$ ; i.e., the names themselves are their Gödel numbers,  $G^{x_{ij}} = x_j^{(i)}$ ). The  $S_{cx0}$  is a set/class that is a



member of nothing (a so-called proper or ultimate class [Quine, 1969]), its cardinality is infinite but countable. Thus, BSDT PL is in general a transfinite language: though its vocabulary is limited at any moment, it may by request arbitrary be enlarged to the extent constrained mainly by animal's morphology only.



**Figure 2.** Explaining the BSDT infinity hypothesis. **A**, Dividing the  $c_{x0}$  into two parts; infinite in the past strings  $c_{xp}$ ,  $c_{xq}$ ,  $c_{xr}$ , thing lines that are the same from their beginning to vertical dashed line 1; prefixes  $x_0^{(p)}$ ,  $x_0^{(q)}$ ,  $x_0^{(r)}$ , thick line segments of lengths  $p$ ,  $q$ ,  $r$  (they are the values of string variables  $x^{(p)}$ ,  $x^{(q)}$ ,  $x^{(r)}$ ); relations between the lengths of  $c_{xp}$  and  $c_{xq}$ ,  $c_{xq}$  and  $c_{xr}$ ,  $c_{xr}$  and  $c_{xp}$  are as follows:  $l(c_{xp}) - l(c_{xq}) = -(p - q) > 0$ ,  $l(c_{xq}) - l(c_{xr}) = -(q - r) < 0$ ,  $l(c_{xr}) - l(c_{xp}) = -(r - p) > 0$  (cf. distances between dashed lines 1 and 2, 1 and 3, 2 and 3). **B**, Graphical presentation of string categories; string variables  $x^{(p)}$ ,  $x^{(q)}$ ,  $x^{(r)}$  (distances between dashed lines 3 and 4, 2 and 3, 1 and 2) are constituents for compound variables  $x^{(qp)}$ ,  $x^{(rqp)}$  (distances between dashed lines 2 and 4, 1 and 4);  $c_{xp}$ ,  $c_{xqp}$ ,  $c_{xrqp}$  are infinite contexts (thing lines) for variables  $x^{(p)}$ ,  $x^{(qp)}$ ,  $x^{(rqp)}$  (thick lines of lengths  $p$ ,  $q + p$ ,  $r + q + p$ ). If to replace strings  $x_0^{(p)}$ ,  $x_0^{(q)}$ ,  $x_0^{(r)}$  by string variables  $x^{(p)}$ ,  $x^{(q)}$ ,  $x^{(r)}$  then lines in **A** represent respective string categories shown in **B**. The string  $c_{x0}$  (or Chaitin's  $\Omega$  [Chaitin, 1998], horizontal lines) begins infinitely far 'in the past' (dashed segments on the left) and ends 'at the present' (vertical dashed lines 4).

### 3. BSDT PL Meaning Functions

For  $c_{xi}x_j^{(i)}$ ,  $x_j^{(i)}$  is the string's unique but meaningless *name*;  $c_{xi}$  is the name's infinite in the past *context*; the name's *meaning* that an animal has *actually* in mind is  $M(x_j^{(i)}) = c_{xi}x_j^{(i)}$ . It is defined by both the name's context,  $c_{xi}$ , and the name itself,  $x_j^{(i)}$ . We refer to an  $M(x_j^{(i)})$ ,  $M(x_j^{(i)}) \in S_{c_{xi}}$ , as a BSDT *meaning* or *mind function* given context,  $c_{xi}$ ; the  $M(x_j^{(i)})$  transforms the names from their meaningless,  $x_j^{(i)}$ , to their meaningful,  $c_{xi}x_j^{(i)}$ , form. The length of the  $M(x_j^{(i)})$  in bits,  $l(M(x_j^{(i)}))$ , is thought of as the name's *meaning complexity* (Sect. 5).

The context,  $c_{xi}$ , is usually shared by a set/range of names,  $x_j^{(i)}$ , that are the values of string variable  $x^{(i)}$ . We say that  $C(x^{(i)}) = c_{xi}x^{(i)}$  is given the  $c_{xi}$  a *category/concept/notion/class* of names  $x_j^{(i)}$  or values of  $x^{(i)}$ ,  $x_j^{(i)} \in S_{xi}$  and  $|S_{xi}| = 2^i$ . The  $i$  is *ensemble complexity*<sup>1</sup> of the  $C(x^{(i)})$  given its context,  $c_{xi}$ . Hence, the complexities of different categories may only be compared if they share the same context. The set of  $|S_{xi}|$  of names  $x_j^{(i)}$  constitute given the  $c_{xi}$  the set of the category's *synonyms/features/attributes/traits/properties*. Thus, a synonymy (and an interchangeability of synonyms in their following use) takes place only among the members,  $x_j^{(i)}$ , of the same category of names,  $C(x^{(i)})$ . That is the BSDT PL's semantic *rule of identity* or an explication of its identity sign/connective '='.

When particular real-world thing/object/event related to the  $C(x^{(i)})$  is named, each the category's item is only specified by the value  $x_j^{(i)}$  of string variable  $x^{(i)}$  or, in other words, by the name's content only – a *randomly* chosen arrangement of its  $i$  positive and negative components. (Here the rigidity and accidantelity of names/designations usually treated as different notions, e.g. [Kripke1990, p. 60], are reconciled.) To name the  $C(x^{(i)})$  given the  $c_{xi}$ ,

<sup>1</sup> It is like information or entropy [Shannon, 1948] for the  $C(x^{(i)})$  given  $c_{xi}$ ,  $H(C(x^{(i)})) = H(x^{(i)})$ . The category  $C(x^{(i)})$  has  $|S_{xi}| = 2^i$  synonyms/states, the probability of occurring of each of them is  $P(x_j^{(i)}) = 1/|S_{xi}|$ . Shannon's information/entropy/complexity is  $H(x^{(i)}) = - \sum (\log_2 P(x_j^{(i)})) / |S_{xi}| = i$  bits;  $j = 1, 2, \dots, 2^i$ .

suffice it to fix the  $i$ . Then the value of  $i$ , if it is written as a string,  $u_s^{(r)}$ , is the *name of the category*.<sup>1</sup> Indeed, knowing the  $i$ ,  $|S_{xi}|$  of names constituting the category can be specified by their contents or, what is the same, by the ordinals<sup>2</sup> 1 to  $|S_{xi}|$  in binary string notations. The category's constituent names,  $x_j^{(i)}$ , are given context nothing more than the ordinal (natural) numbers with their upper limit  $|S_{xi}|$ . In other words, any category as a whole,  $C(x^{(i)})$ , can completely be specified by its context,  $c_{xi}$ , and by either all the constituting it synonyms (the values of  $x^{(i)}$  ranged in  $S_{xi}$ ) or the category's unique name,  $u_s^{(r)}$ . The meaning of the category's name is defined by both the name itself,  $u_s^{(r)}$ , and the name's context,  $c_{ur}$ , i.e.  $M(u_s^{(r)}) = c_{ur}u_s^{(r)}$  under condition  $c_{xi} = c_{ur}$ .

Meanings  $M(u_s^{(r)})$ ,  $M(u_s^{(r)}) \in S_{cu0}$ , and  $M(x_j^{(i)})$ ,  $M(x_j^{(i)}) \in S_{cx0}$ , are members of sets  $S_{cu0}$  and  $S_{cx0}$  that comprise the strings that share their infinite initial part but differ in length, by  $i - r$  bits (prefixes  $u_s^{(r)}$  and  $x_j^{(i)}$  provide Gödel numbering of members of  $S_{cu0}$  and  $S_{cx0}$ , respectively). Names  $x_j^{(i)}$  and  $u_s^{(r)}$  are of different levels<sup>2</sup> and denote things of similar but different nature (with mostly common but different in length evo-devo prehistories, see below). Strings  $M(u_s^{(r)}) = c_{ur}u_s^{(r)}$  may be treated as fractions of strings  $M(x_j^{(i)}) = c_{xi}x_j^{(i)}$  if strings  $u_s^{(r)}$  are treated as *fringe* (left-most in Fig. 2B and Sect. 4) fractions of compound strings  $x_j^{(i)}$ . In such cases, the meaning of  $x_j^{(i)}$  is precise and clear while the meaning of  $u_s^{(r)}$  is vague. Where  $u_s^{(r)}$  is treated as a fringe of  $x_j^{(i)}$ , the vagueness of meaning of the  $u_s^{(r)}$  is measured by its colored  $2^{i-r}$ -nonlocality (Sect. 2). Thus, in just described 'fuzzy' or 'vague' sense only, it may be that  $S_{cu0} \subseteq S_{cx0}$  and meanings  $c_{xi}x_j^{(i)}$  and  $c_{ur}u_s^{(r)}$  are related given the  $c_{xi} = c_{ur}$ .

Given the  $c_{xi}$  strings  $x_j^{(i)}$  are perfect for most profitable, concise (the best, incompressible) binary coding of distinctions in naming the category's items. Keeping the sum,  $l(c_{x0}) = l(c_{xi}) + i$ , but varying the length of  $x_j^{(i)}$ ,  $i$ , an infinite series of infinite and common in the past strings  $c_{xi}x_j^{(i)}$ , with different in length right-most *random-valued* fractions  $x_j^{(i)}$ , can be obtained in a way when each longer  $x_j^{(i)}$  contains as its right-most part any shorter one. Such infinite strings,  $c_{xi}x_j^{(i)} \in S_{cx0}$ , are by definition [Chaitin, 1998] *random* and *uncomputable*.

Among infinite binary strings with definite but uncomputable (random) components, it is selected the one,  $\Omega$  [Chaitin, 1998], that given computer (its hardware and software) provides halting probabilities for algorithms of any length; namely  $i$  right-most bits of  $\Omega$ ,  $\Omega_i$ , give halting probabilities for algorithms not longer than  $i$ . Hence, any  $c_{xi}x_j^{(i)} \in S_{cx0}$  can be interpreted as the  $\Omega$  whose the  $ij$ th right-most fraction of  $i$  bits,  $\Omega_{ij}$ , is specified given the  $ij$ th randomly chosen self-delimiting computer. Moreover, random strings  $x_j^{(i)}$  can be treated [Chaitin, 1998] as true, unprovable assertions or irreducible mathematical facts that can only be deduced by adding them as axioms. The  $|S_{cx0}|$  provides the amount of different possible finitely defined computers self-delimiting by respective values of  $i$ . Since meaning functions are the items of  $S_{cx0}$ ,  $M(x_j^{(i)}) \in S_{cx0}$ , they are also random and uncomputable, and their arguments,  $x_j^{(i)}$ , are, given the  $ij$ th computer, the  $ij$ th fractions of  $\Omega$ ,  $\Omega_{ij} = x_j^{(i)}$ ; the properties of  $c_{xi}$ , except its property of being infinitely long and common in the past for different  $x_j^{(i)}$ , are here not essential in general.

Each string  $x_j^{(i)}$  is selected by its BSDT abstract selectional machine,  $ASM(x_j^{(i)})$  [Gopych, 2007], designed beforehand to do this. Explicit meaning of  $x_j^{(i)}$  is presented as infinite-in-length *symbolically written* complete evo-devo (evolutionary and developmental) prehistory,  $c_{xi}x_j^{(i)}$ , of designing the real-world  $ASM(x_j^{(i)})$ . Since an animal's real-world *physical*  $ASM(x_j^{(i)})$ , i.e. implicit meaning of  $x_j^{(i)}$ , is in turn the  $ij$ th network internal ('mental') representation of the  $ij$ th internal/external object/event [Gopych, 2008b, 2009b], the meaning of  $x_j^{(i)}$  depends on the current animal state and animal relations to the environment. In that sense, the name's meaning is just the

<sup>1</sup> The name  $u_s^{(r)}$  is the  $s$ th value of binary string variable  $u^{(r)}$  or a vector of the dimension  $r$ . The  $r$  equals either  $\log_2(N_{xi})$  if it is a natural number or  $\lceil \log_2(N_{xi}) \rceil + 1$  in opposite case, [...] means a part of a real number before its decimal point and  $i > 0$ ;  $u_s^{(r)} \in S_{ur}$ ,  $|S_{ur}| = 2^r$ ;  $l(c_{xi}x_j^{(i)}) - l(c_{ur}u_s^{(r)}) = i - r > 0$ . Hence, the names  $x_j^{(i)}$  are longer (more complex) than the names  $u_s^{(r)}$ . The form  $C(u^{(r)}) = c_{ur}u^{(r)}$  defines, we say, a *higher-level* category whose *higher-level* constituent names,  $u_s^{(r)}$ , designate given the context,  $c_{ur}$ ,  $|S_{ur}|$  of different *first-level* categories. One of such categories,  $C(x^{(i)})$ , of *first-level* constituent names,  $x_j^{(i)}$ , is discussed. Other first-level categories related the  $C(u^{(r)})$  and more higher-level categories are here not considered. An infinitely deep hierarchy of property names and related notion names built in such a way may be treated as their BSDT LT *ontology*.

<sup>2</sup> The possibility 'to take the ordinals themselves as primitive terms' was still envisaged in [Gödel, 1946].

animal's being in a certain *psychological* state. (Note, for any binary message  $x_j^{(i)}$ , its meaning,  $M(x_j^{(i)})$ , exists in its implicit,  $M_{\text{impl}}(x_j^{(i)})$ , or explicit,  $M_{\text{expl}}(x_j^{(i)})$ , form [Gopych, 2009b]. Since this paper is mainly devoted to a *symbolic* language, we are here usually dealing with  $M(x_j^{(i)}) = M_{\text{expl}}(x_j^{(i)})$ .)

The context  $c_{x_i}$  gives the description of designing the  $C(x^{(i)})$ , a category of ASMs selecting the values of  $x^{(i)}$ ,  $x_j^{(i)} \in S_{x_i}$ . Thus, given context the meaning of  $C(x^{(i)})$  is  $|S_{x_i}|$  of real-world physical ASMs,  $\text{ASM}(x_j^{(i)})$  with  $j$  from 1 to  $|S_{x_i}|$ , representing respective things. In other words, the meaning of a category of names is eventually a set of real-world objects/events given to an animal *through* its particular psychological state – the state of activity of complex hierarchies (neural subspaces) of typical neural networks constituting a common milieu for signal processing, memory storage/retrieval, decision-making [Gopych, 2008b] and consciousness [Gopych, 2009b]. Resulting meanings are used for animal mental (brain) *meaningfullsemantic* computations needed for animal adaptive behavior in its natural (including social) environment. This explains why BSDT PL names taken in isolation are actually *meaningless* by themselves: their meanings are mostly in their contexts that can be written as infinite binary strings and implemented as real-world ASM hierarchies (neural subspaces) playing the role of separate computers devoted to serve (to generate the meaning for) respective meaningless names. Hence, the BSDT predicts the following effect. By examining in experiment a given ASM hierarchy that generates the meaning of particular  $x_j^{(i)}$ , parameters describing the hierarchy itself may successfully be found but the content of  $x_j^{(i)}$  – specific arrangement of its components – will always remain unknown (it becomes irrelevant and not essential once the hierarchy in question was picked out for the investigation).

#### 4. Meanings of BSDT PL Compound Names

From the comparing of Fig. 2A and B follows that, for a compound string variable, e.g.  $x^{(rap)} = x^{(i)}x^{(a)}x^{(p)}$ , only its right-most constituent,  $x^{(p)}$ , can correctly be assigned to its context,  $c_{x_p}$ , and can turn out to be a category of names,  $c_{x_p}x^{(p)}$ , with constituent names  $x_j^{(p)}$  ranged in  $S_{x_p}$ . Other constituents of the  $x^{(rap)}$ , though in it they are certainly presented, cannot manifest their precise meanings and remain to an extent fundamentally vague, opaque, or oblique. We refer to the compound string's right-most variable (category of names,  $c_{x_p}x^{(p)}$ ) as a focal one, the other variables (categories) are referred to as its surround or '*fringe*'. One of  $|S_{x_p}|$  of values of  $x^{(p)}$ , e.g.  $x_0^{(p)}$ , can be transformed into the meaningful string,  $c_{x_p}x_0^{(p)}$ , whereas all its other values,  $x_j^{(p)} \neq x_0^{(p)}$ , are in  $x^{(p)}$  and  $x^{(rap)}$  only *implicitly* presented, as available in principle possibilities. The right-most compound constituent of  $x^{(rap)}$ ,  $x^{(ap)} = x^{(a)}x^{(p)}$ , can in turn be interpreted as a category of names,  $c_{x_{ap}}x^{(ap)}$ , with its precisely defined context,  $c_{x_{ap}}$ ; one of  $|S_{x_{ap}}|$  of values of the  $x^{(ap)}$ , e.g.  $x_0^{(ap)}$ , can in turn be transformed into a meaningful string  $c_{x_{ap}}x_0^{(ap)}$ , whereas all its other values,  $x_j^{(ap)} \neq x_0^{(ap)}$ , are in  $x^{(ap)}$  and  $x^{(rap)}$  only implicitly presented. Finally, the compound variable  $x^{(rap)}$  can in itself be presented as a category of names,  $c_{x_{rap}}x^{(rap)}$ , with its precisely defined context,  $c_{x_{rap}}$ ; one of  $|S_{x_{rap}}|$  of values of the  $x^{(rap)}$ , e.g.  $x_0^{(rap)}$ , may in turn be transformed into a meaningful string  $c_{x_{rap}}x_0^{(rap)}$ , whereas all its other values,  $x_j^{(rap)} \neq x_0^{(rap)}$ , are in  $x^{(rap)}$  only implicitly presented. Because of laws of communicativity and associativity, any number of compound string's constituents may be arranged in an arbitrary fashion and create a compound *substring* that, as any other compound string's constituent, can take in it the right-most (focal) position. We see, compound string's meaning may be constructed of meanings of its constituents either 'analytically' (taking it serially, 'term by term' [Quine, 1992], as a set of meanings of its constituents) or 'holophrasically' (taking it as 'a seamless whole' [Quine, 1992], not noticing explicitly the meanings of compound string's constituents). In any case, the BSDT PL lacks direct 'compositional semantics'.

Compound names are thought of as BSDT PL *sentences* and, simultaneously, BSDT PL descriptions of BSDT neural subspaces – the neural network hierarchical milieu (cf. Fig. 3 in [Gopych, 2008b] and Fig. 1 in [Gopych, 2009b]) for BSDT signal processing, memory, decision-making and consciousness [Gopych, 2008b, 2009b]. If so, the number of compound name's constituents is also treated as an animal's logical/reasoning deepness [Gopych, 2009b], the value of focal string variable corresponds to memory's feature/attribute that is currently in the focus of attention, the focal variable's fringe is the fringe of animal's memory or consciousness. Compound name's 'holophrasical' presentation describes the perception/remembering of a scene in a whole (diffuse focus of attention), whereas its 'analytical' presentation describes the serial perception/remembering of the scene by its

(compound or not) fractions (acute focus of attention). Any paraphrase (rearrangement of constituents) of the BSDT sentence cannot change its meanings as a whole. The combining of BSDT PL names into sentences and operating on them should always be performed *given* their meanings. In that sense BSDT PL semantics (interpreting the names) is primary with respect to its syntax (rules for the construction and transformation of sentences) though they are of course closely interrelated.

## 5. BSDT PL Meaning Complexity

An  $x_j^{(i)}$  becomes meaningful once it is attached to its respective infinite in the past context,  $c_{xi}$ , describing the ASM devoted to select the  $x_j^{(i)}$ ,  $ASM(x_j^{(i)})$  [Gopych, 2007]. For all the first-level names, resulting strings,  $M(x_j^{(i)}) = c_{xi}x_j^{(i)}$ , have the same total length or *meaning complexity*,  $I(c_{xi}x_j^{(i)}) = I(c_{xi}) + i = I(c_{x0})$ , reflecting in a sense animal's complexity, *not* the complexities of things the names designate (note, all animal's tissues share their genetic structure). For a higher-level name,  $u_s^{(r)}$ , its meaning complexity,  $I(c_{ur}u_s^{(r)}) = I(c_{ur}) + r = I(c_{u0})$ , is described in the same way though it is smaller by  $i - r$  bits than meaning complexity,  $I(c_{x0})$ , of its first-level names,  $x^{(i)}$  (footnote 2). Thus, meaning complexities of names of real-world things of any nature or notions of them should simultaneously be specified by *two* kinds of additive complexities: infinite *context complexity*, e.g.  $I(c_{xi})$  for given-level meaningless names  $x_j^{(i)}$ , and their finite *ensemble complexity*,  $i$ . Ensemble complexity or Shannon information/entropy (footnote 1) is defined for a set of  $|S_{xi}|$  of names and quantifies average or *ensemble properties* of meaningless names,  $x_j^{(i)}$ , while the ensemble's *individuality* is quantified by their common context complexity,  $I(c_{xi})$ . Infinite meaning complexity quantifies the *individuality* or the sameness of (meanings of) either a separate name or a separate category of names (or the sameness of animal's respective psychological states).

A name's meaning complexity or the first Cantor's cardinal number,  $I(c_{xi}x_j^{(i)}) = I(c_{x0})$ , is infinite but, because of our infinity hypothesis (Sect. 2), comparable with meaning complexities of other names. Any category of compound names,  $c_{xqp}x^{(qp)}$ , with respect to its constituent categories,  $c_{xp}x^{(p)}$  and  $c_{xq}x^{(q)}$ , has smaller context complexity, i.e.  $I(c_{xqp}) < I(c_{xp})$  and  $I(c_{xqp}) < I(c_{xq})$ , and larger ensemble complexity, i.e.  $q + p > p$  and  $q + p > q$ , whereas their constituent names, no matter whether they are compound or not, have the same meaning complexity,  $I(c_{x0})$ , and no ensemble complexity (see Fig. 2). In other words, for names of the same level, any association of them cannot change their meanings acquired beforehand or make the compound name more meaningful than its constituents. Only the names of different levels have different meaning complexities.

## 6. BSDT PL Truths, Communication Paradox

The  $j$ th name  $x_j^{(i)}$  of the length  $i$  is true if its meaning,  $M(x_j^{(i)})$ , is true or, in other words, strings  $c_{xi}$  and  $x_j^{(i)}$  (the name's context and the name itself) are correctly adjoined to or satisfy each other. Thus, for all the category's names,  $x_j^{(i)} \in S_{xi}$ , their meanings,  $M(x_j^{(i)}) = c_{xi}x_j^{(i)}$ , are to be *consistent* or, in other words, their context string,  $c_{xi}$ , has to be the same and the relation  $I(c_{xi}) + i = I(c_{x0})$  has to be valid. In opposite case, the  $j$ th meaningful name is false. Since the cardinality of  $S_{cx0}$ ,  $|S_{cx0}|$ , is infinite, the number of BSDT PL truths (true meaningful names) is infinite in general and, for any such meaningful string, its truth-value (either 'true' or 'false') certainly exists. At the same time, because compound string's fringe constituents are within it colored due to their discrete nonlocality (Sect. 2), the meanings of fringes are neither true nor false; they are 'in limbo' [Quine, 1992] or fuzzy/vague in meaning. This is a BSDT PL counterpart to Gödel's incompleteness theorem (note, particular axioms and theorems for which Gödel's results hold are in our terms an infinite fraction of infinite in number meaningless strings  $x_j^{(i)}$ ). The BSDT PL truths are here introduced as a correspondence between names and reality. All the meaningful names, e.g.  $c_{xi}x_j^{(i)} \in S_{cx0}$ , are by definition true because they name particular real-world things given to an animal through its respective psychological states. Thus, for the BSDT PL, the truth is a norm and the falsity is an anomaly caused, e.g., by animal's disfunction or disease. In any case, there is *no* lie and *no* liar paradox, a severe problem of formal logic, e.g. [Quine, 1992], and a source of the Gödel's incompleteness which does not hold for BSDT PL *meaningful* names,  $c_{xi}x_j^{(i)}$ . This inference is caused by the fact that BSDT PL name meanings are the ones that animals/humans have actually in mind and are mainly concentrated in their infinite contexts.

Since complete symbolic meaningful descriptions of BSDT PL names,  $c_{xi}x_j^{(i)}$ , are fundamentally *infinite*, they cannot be communicated even in principle. To solve this *communication paradox* [Gopych, 2009b], we appeal to our infinity hypothesis. Let an ASM-sender and an ASM-receiver share their prehistory/context, i.e. let they were designed beforehand to perform the same meaningful function – selecting a finite-in-length symbolic message,  $x_j^{(i)}$ , given its infinite context,  $c_{xi}$ . Only in such a case the meaning of  $x_j^{(i)}$ ,  $c_{xi}x_j^{(i)}$ , is equally coded/decoded/interpreted or, in other words, equally *understood* by both parties. For this reason, and because the name's meaning is the psychological state an animal experiences when it produces or perceives the name (see Sect. 3), in meaningful information exchange, both the sender and the receiver are to be *physically* and *functionally* equivalent. Thus, correct understanding of meanings of different names is only possible if the numbers of ASM-senders, of ASM-receivers and of names to be produced/received by particular animal are equal to each other. Of this follows that the paradigm of network coding/decoding of meaningful messages should be 'one-memory-trace-per-one-network', the rule actually employed by the BSDT [Gopych, 2008b]. According to recent empirical neuroscience findings (discovering the coding by 'synapse assemblies' [Xu, 2009; Yang, 2009]), the same coding/decoding rule is also to be used by animals. Since ASM prehistories include complete stories of ASM *individual* development, completely equivalent ASMs/ASM hierarchies are practically impossible. For this reason, meaning understanding is always either to some extent vague or even wrong in worse cases. The sameness of ASM contexts is here the key. An animal involuntarily demonstrates its actual internal psychological states (meanings of respective BSDT PL names, Sect. 3) by its basic/principal/involuntary behaviors. Another animal of the same species watching the first one would be able to understand it, because they the both have similar bodies and nerve systems (i.e. the same morphology produced by their common evo-devo prehistories) and would similar behave in similar environmental situations. For animals, to understand each other's behavior means to match the activities of respective parts of their nerve systems (neural subspaces or ASM hierarchies representing the respective names) no matter whether particular animal behaves or only watches the behavior of others. This BSDT PL prediction is completely consistent with well known neuroscience finding of 'mirror' neurons – the ones that are active in motor brain areas of animals that are only watching respective motor behavior of others, e.g. [Rizzolatti, 2004]. For meaningful interanimal behavioral information exchange (involuntary understanding of each other's basic behaviors), there is no need in any specially developed additional symbolic communication system (language). It may only be required to *enrich* animal's basic communication faculties.

The truth values as such are together with names never communicated. They should always be discovered and confirmed by checking the correspondence of names to the reality or, more directly, to animal's respective psychological states. Most probably it is doing by physical/anatomical segregation and specification of communication channels (input/output sensory submodalities) or carriers for communication signals, with the following convergence of relevant channel-specific information. By means of such segregation and specification, particular neural subspaces/ASM hierarchies (or computers for particular mental computations, Sect. 3) are eventually allocated. By the convergence of information from different communication channels, an integral and, consequently, most reliable estimation of the current state of animal's environment has to be achieved.

---

## 7. Some BSDT PL Applications

The semantics proposed generates a manual for practical semantic computations. On the one hand, it is surprisingly simple because it recommends to *have* before the beginning of computations on names  $x_j^{(i)}$  their already known infinite contexts,  $c_{xi}$ , i.e. to use already completely established formal models of reality. Numerous computations in mathematical physics, chemistry, biology, etc could be considered as approximate examples. On the other hand, it is surprisingly complex because it recommends to *perform* beforehand complete symbolic descriptions of infinite contexts,  $c_{xi}$ . This goal can of course never be achieved. That is, the main problem of symbolic semantic computations is the incompleteness of contexts or, in other words, 'mathematical' models of reality. Once they became known (at least approximately), BSDT PL semantic computations are reduced to usual Turing-machine operations on strings  $x_j^{(i)}$  and could easily be performed, e.g. [Gopych, 2006, 2008b].

BSDT decoding algorithms [Gopych, 2008a, 2008b] serve as BSDT PL rules of inference. They are conventional operations on binary vectors  $x_j^{(i)}$ , can be presented as ASMs [Gopych, 2007] and exist in neural network, convolutional and Hamming distance forms that are functionally equivalent and the best [Gopych, 2008a]. To be able to be meaningful, each message  $x_j^{(i)}$  is stored and processed by its separate ASM. At the same time, the BSDT PL says nothing of mechanisms of arrangement/rearrangement of its compound names (sentences) and of selecting their those fragments that are to be placed into the current focus of attention (see Fig. 2B). For this reason, an appeal to a kind of analog (e.g. wave-like) computations seems to be unavoidable (see also [Gopych, 2009a]). An animal's morphology is also essential for verifying the truth-values of names to be communicated. Thus, the BSDT PL as such is *incomplete* and, consequently, *insufficient* to ensure its own running in full.

One of distinctive features of the BSDT PL is that its names are in the norm true (Sect. 6). That is why it is so well to serve as primary language for maintaining an animal's ongoing internal activity. For the same reason, it can serve as a 'source language' whose names (animal's respective psychological states) may next be translated into vocal, gestical, etc tokens of a more elaborate symbolic communication system needed to support information exchange between animals of a group. The more complicate the group's sociality, the more complicate communication system is required to support it (and vice versa). Since among other animals humans do have most complicate sociality, human natural languages are to be most complicate and elaborate. The BSDT PL may be used for the construction of such 'secondary' [von Neumann, 1958] languages whose capacities may be at the level of human natural languages. If so, semantics and syntax of natural languages should be in a broad sense the functions of semantics and syntax of the BSDT PL. Specific form of these functions should be defined by mechanisms of (and innate brain structures for) the translation of the primary language into a secondary one. In that sense the BSDT PL is a precursor for or a counterpart to so-called 'universal grammars', e.g. [Chomsky, 1997]. The BSDT PL is a universal language for communicating the meanings of its words/names or animal's respective psychological states by means of animal's basic/involuntary behaviors that could represent the behavioristic part of its more complex adaptive behaviors. For animals of the same and, in many cases, of different species, thanks to their mirror neurons [Risolatti, 2004], it is intelligible without any efforts. For animals with most primitive sociality or for their *artificial* counterparts, a version of the BSDT PL may serve as an *exhaustive* (though *incomplete*, see above) set of tools for their routine communications.

---

## 8. Discussion

---

Perhaps most important tenet of the BSDT PL is its *infinity hypothesis* – the idea that BSDT contexts specifying the symbolic descriptions of meanings of names are infinite 'in the past' binary strings sharing their infinite initial part (Sect. 2). It is grounded on to date rather established natural phenomenon of the unity (common coevolution) of universe, life, and mind [Gopych, 2009b] and further elaborates, e.g., famous Zermelo's infinity axiom, e.g. [Quine, 1969]. Armed with its infinity hypothesis, the BSDT PL becomes the theory spreading *beyond* the scope of ubiquitous axiomatic method because its main objects (meaningful names and prehistories of ASMs devoted to select them) are inherently infinite and cannot be described taking traditional (Hilbert's Program) finite standpoint.

The duality of complete BSDT PL names,  $c_{xi}x_j^{(i)}$ , presented simultaneously as the name's meaningless random numerical identifier,  $x_j^{(i)}$ , and the name's meaningful context,  $c_{xi}$ , is rather similar to the duality introduced by Frege [Frege, 1892] to distinct the sense from the meaning of a word/sentence of a natural language. A proper name 'has as its meaning a definite object' while the name's sense 'serves to illuminate only a single aspect of the thing meant' [Frege, 1892]. The BSDT PL's name meaning, similar to Frege's, is a real-world thing given to an animal through its respective psychological state (Sect. 3), an analogue of Frege's notion of the 'idea' [Frege, 1892]. Thus, BSDT PL name meaning is actually a real-brain *physical* implementation of respective ASM hierarchy/neural subspace representing a real-world thing. This inference almost completely coincides with the 'direct reference' view popular in philosophy – 'A name means an object. The object is its meaning' [Wittgenstein, 1922; 3.203]. The name's explicit meaning,  $M_{\text{expl}}(x_j^{(i)})$ , is its complete *symbolic* description,  $c_{xi}x_j^{(i)}$ , though this description is of course not the same as its *physical* implementation – particular real-brain neural subspace or the name's implicit meaning,  $M_{\text{impl}}(x_j^{(i)})$  ([Gopych, 2009b], i.e. the BSDT PL reconciles two distinct roles of a name and the

term ‘definition’ – to refer to anything nameable and to designate its description). *Context principle* – ‘never to ask for the meaning of a word in isolation, but only in the context of a proposition’ [Frege, 1884] – is also completely consistent with the BSDT PL. (Note, Frege’s German original *Bedeutung* [Frege, 1892] is translated into English as reference, meaning, denotation, signification, indication, designatum, nominatum, etc. Moreover, Frege himself used “the term ‘sense’ in two senses” – as a description of the name’s meaning and as ‘the way its reference is determined’. ‘Usually Fregean sense is now interpreted as the meaning’ [Kripke, 1990; p. 59, footnote 22]. Since meanings of BSDT PL names are indeed used in most general sense, we prefer the term ‘meaning’.)

On the other hand, the duality of BSDT PL names radically differs from Fregean or neo-Fregean duality. The infinity of meaningful BSDT PL names,  $c_{xi}x_j^{(i)}$ , and our infinity hypothesis are the reasons. Taken separately, BSDT PL names,  $x_j^{(i)}$ , are completely meaningless and have neither the ‘sense’ nor the ‘meaning’. They are the simplest conceivable numerical identifiers,  $G^{x_{ij}} = x_j^{(i)}$  (Sect. 2), providing Gödel numbering of all the possible BSDT PL meaningful names/sentences/expressions,  $c_{xi}x_j^{(i)} \in S_{cx0}$ . BSDT PL Gödel numbers,  $G^{x_{ij}}$ , are written in binary string notations and represent the ordinals counting in a fixed but arbitrary chosen order appropriate real-world things whose properties are given to animals as *axiomatic* truths. Inherent randomness and uncomputability of BSDT PL names/Gödel numbers admit them to be interpreted as *ij*th fractions of Chaitin’s  $\Omega$ ,  $x_j^{(i)} = G^{x_{ij}} = \Omega^{x_{ij}}$  (Sect. 3). Among infinite number of BSDT PL string names,  $x_j^{(i)}$ , because the set of them is complete (includes all possible variants) there are always the ones that in a system of notations code finite-in-length axioms, inference rules, and theorems of any axiomatic calculus. At the same time, we cannot guarantee that particular calculus’ (e.g. Whighthead and Russell’s [Russell, 1903]) axioms, inference rules and theorems, though infinite in number, exhaust the BSDT PL names. Indeed, according to Gödel’s incompleteness theorem, that is not the case. Moreover, any attempt to syntactically attach to a meaningless string a meaning additional to its inherent one of being a number leads to contradictions, as in the case of famous liar (or Berry’s, cf. [Boolos, 1989]) paradox. Since BSDT PL meaningful names are always true, for them there is *no* liar paradox, *no* related contradictions, and *no* incompleteness in the sense of Gödel. Instead, the problem arises of vague relations between meanings of names of different levels (Sect. 3 and footnote 2). The root of this problem is in colored nonlocality of higher-level names,  $u_s^{(i)}$ , and our fundamental inability to strictly bound the meanings ( $c_{ur}u_s^{(i)}$  and  $c_{xi}x_j^{(i)}$ ,  $c_{ur} = c_{xi}$ ) of names/Gödel numbers ( $G^{u_{rs}} = u_s^{(i)}$  and  $G^{x_{ij}} = x_j^{(i)}$ ) having different meaning complexities,  $I(c_{ur}u_s^{(i)}) \neq I(c_{xi}x_j^{(i)})$ , and related to different ultimate classes, namely  $c_{ur}u_s^{(i)} \in S_{cu0}$ ,  $c_{xi}x_j^{(i)} \in S_{cx0}$  assuming  $c_{ur} = c_{xi}$  (Sects. 3, 4).

BSDT PL notion of meaning complexity (Sect. 5) embraces the notions of information/entropy [Shannon, 1948] and Kolmogorov [Li, 1997] or algorithmic [Chaitin, 1998] complexity. Meaning complexity integrates the descriptions of ensemble properties and individual properties of things and provide a possibility of their quantitative comparing. Given the  $c_{xi}$ , since  $x_j^{(i)} = G^{x_{ij}} = \Omega^{x_{ij}}$ , the mentioned complexities for BSDT PL names,  $x_j^{(i)}$ , respective Gödel numbers,  $G^{x_{ij}}$ , and respective halting probabilities,  $\Omega^{x_{ij}}$ , coincide with each other. Gell-Mann and Lloyd [Gell-Mann, 1996] combined in a finite manner Kolmogorov information and Shannon information/entropy to introduce an *effective complexity* of an ensemble in a string form – that is a loose counterpart to BSDT PL context complexity – and combined effective complexity with Shannon information/entropy to obtain total information over an ensemble – that is a loose counterpart to BSDT PL meaning complexity. For BSDT PL meaning complexity, it is its infiniteness that is given context its decisive distinction.

The BSDT PL convention on truth (Sect. 6) essentially differs from Tarski’s *convention T* [Tarski, 1935]. Tarski’s definition is *syntactical* and holds for a given axiomatically defined pair object-language/metalanguage only, whereas BSDT PL definition is *semantical* and uses the reality for checking the truths. Current truth-values of BSDT PL names are *unique* and *conclusive*, any hierarchy of them is neither possible nor required, in contrast to Tarski’s syntactical approach implying that for any metalanguage its meta-metalanguage can in turn be conceived and so up. For this reason, for each ‘higher-level’ pair metalanguage/meta-metalanguage its higher-level truth (the truth of sentences of respective higher-level metalanguage) could in general be defined. Tarski’s truths are *relative*, BSDT PL truths are *absolute* though sometimes ambiguous. An ambiguity arises for meanings (truths) of names of different levels – it is caused by the difference in their meaning complexities or colored nonlocalities of higher-level/fringe names. For names of a given level, their truths are always unambiguous and

defined in the same way, as in Sect. 6. Name hierarchies (ontologies) and ambiguities between meanings (truths) of names of different levels are implemented (in the brain) by means of BSDT neural subspaces/ASM hierarchies for signal processing, memory, decision-making and consciousness [Gopych, 2008b, 2009b].

---

## 9. Conclusion

Instead of explicit (Fregean) use of the calculus of propositions, the BSDT PL employs the calculus of finite-in-length strings/vectors perfectly tuned to each other by their common infinite initial parts (contexts). Since strings/vectors/names participated in BSDT PL inference rules (BSDT ASMs, Sect. 7) are finite, calculations on them can given the context be performed either exactly or, given any accuracy, approximately, e.g. [Gopych, 2006, 2008b]. In that sense BSDT PL predictions are strict and definite. On the other hand, since BSDT PL names/sentences are always to be attached to their infinite contexts, context's accuracy becomes crucial for the understanding of respective computations. For this reason, BSDT PL predictions are as accurate and definite as accurate and definite are infinite contexts supplied (estimated) by methods beyond the BSDT PL.

The BSDT PL above introduced is completely consistent with contemporary science and provides predictions (Sect. 3) testable in the experiment. It explains basic/involuntary behaviors of animals of simplest sociality (Sect. 6) and, consequently, gives the rules (though they are fundamentally incomplete) for designing their *approximate* copies, as close to their originals as possible. Practical semantic computations and designing the languages with capacities approaching natural languages are among other perspectives of BSDT PL possible applications.

---

## Acknowledgements

I am grateful to my family and my friends for their help and support. The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria [www.ithea.org](http://www.ithea.org), and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine [www.aduis.com.ua](http://www.aduis.com.ua).

---

## Bibliography

- [Boolos, 1989] G. Boolos. A New Proof of the Gödel Incompleteness Theory. *AMS Notices* 36, 388-390, 1989.
- [Chaitin, 1998] G. Chaitin. *The Limits of Mathematics*. Springer-Verlag, Singapore, 1998.
- [Chomsky, 1997] N. Chomsky. *The Minimalist Program*. MIT Press, Cambridge, MA, 1997.
- [Frege, 1884] G. Frege. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Tr. J.L. Austin. Blackwell, Oxford, 1884/1959.
- [Frege, 1892] G. Frege. On Sense and Meaning. In: *Philosophy for the 21st Century*, pp. 506-511. Ed. S.M. Cahn. Oxford University Press, Oxford, 1892/2003.
- [Gell-Mann, 1996] M. Gell-Mann, S. Lloyd. Information Measures, Effective Complexity, and Total Information. *Complexity* 2(1), 44-52, 1996.
- [Gopych, 2006] P.M. Gopych. Performance of BSDT Decoding Algorithms Based on Locally Damaged Neural Networks. *LNCS*, vol. 4224, pp. 199-206. Springer-Verlag, Berlin-Heidelberg, 2006.
- [Gopych, 2007] P. Gopych. Minimal BSDT Abstract Selectional Machines and their Selectional and Computational Performance. *LNCS*, vol. 4881, pp. 198-208. Springer, Berlin-Heidelberg, 2007.
- [Gopych, 2008a] P.M. Gopych. Elements of the Binary Signal Detection Theory, BSDT. In: *New Research in Neural Networks*, pp. 55-63. Eds. M. Yoshida, H. Sato. Nova Science, New York, 2008.
- [Gopych, 2008b] P. Gopych. Biologically Plausible BSDT Recognition of Complex Images: The Case of Human Faces. *Int. J. Neural Systems* 18, 527-545, 2008.
- [Gopych, 2009a] P. Gopych. BSDT Multi-valued Coding in Discrete Spaces. *ASC*, vol. 53, pp. 258-265. Springer, Berlin-Heidelberg, 2009.
- [Gopych, 2009b] P. Gopych. BSDT Atom of Consciousness Model: The Unity and Modularity of Consciousness. *LNCS*, vol. 5769, pp. 54-64. Springer, Berlin-Heidelberg, 2009.



- 
- [Gödel, 1946] K.Gödel. Remarks Before Princeton Bicentennial Conference on Problems of Mathematics. In: Ed. M.Davis The Undecidable, pp. 84-88. Raven Press, New York, 1946/1985.
- [Kripke, 1990] S.Kripke. Naming and Necessity, rvd. ed. Basil Blackwell, Oxford, 1990.
- [Li, 1997] M.Li, P.Vitanyi. An Introduction to Kolmogorov Complexity and its Applications, 2nd ed. Springer, Berlin-Heidelberg, 1997.
- [Quine, 1969] W.V.Quine Set Theory and its Logic. Harvard University Press, Cambridge, MA, 1969.
- [Quine, 1992] W.V.Quine. Pursuit of Truth. Harvard University Press, Cambridge, MA, 1992.
- [Rizzolatti, 2004] G.Rizzolatti, L.Craighero. The Mirror-neuron System. Ann. Rev. Neurosci., 27, 169-192, 2004.
- [Russell, 1903] B.Russell. Principles of Mathematics. Routledge, London-New York, 1903/2010.
- [Searle, 1981] J.R.Searle. Expression and Meaning. Cambridge University Press, Cambridge, MA, 1981.
- [Shannon, 1948] C.Shannon. A Mathematical Theory of Communication. Bell Syst. Techn. J. 27, 379-423, 623-656, 1948.
- [Tarski, 1935] A.Tarski. Logic, Semantics, Metamathematics, 2nd ed. Oxford University Press, Oxford, 1935/1983.
- [von Neumann, 1958] J. von Neumann. The Computer and the Brain. Yale University Press, New Haven, 1958/1986.
- [Wittgenstein, 1922] L.Wittgenstein. Tractatus Logico-philosophicus. Routledge, London-New York, 1922/1974.
- [Xu, 2009] T.Xu, X.Yu, A.J.Perlik, W.F.Tobin, J.A.Zweig, K.Tennant, T.Jones, Y.Zuo. Rapid Formation and Selective Stabilization of Synapses for Enduring Motor Memories. Nature 462 (7275), 915-919, 2009.
- [Yang, 2009] G.Yang, F.Pan, W.-B.Gan. Stably Maintained Dendritic Spines are Associated with Lifelong Memories. Nature 462 (7275), 920-924, 2009.

---

### Authors' Information

---



**Petro Gopych** – Universal Power Systems USA-Ukraine LLC, Science adviser, 3 Kotsarskaya Street, Kharkiv 61012 Ukraine; e-mail: [pmgopych@gmail.com](mailto:pmgopych@gmail.com), [pmg@kharkov.com](mailto:pmg@kharkov.com)  
Major Fields of Scientific Research: Neural networks and their applications, Human performance, Consciousness

## Neural Networks and Pattern Recognition

### СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ АРХИТЕКТУРЫ НЕЙРОСЕТЕВЫХ КЛАССИФИКАТОРОВ

Альберт Воронин, Юрий Зиятдинов, Анна Антонюк

**Аннотация.** Рассматривается метод оптимизации архитектуры нейросетевого классификатора: поисковый метод симплекс-планирования. С помощью программы обучения нейронной сети и программы оптимизации проводится исследование корректности выбранного поискового метода для выявления наилучшей архитектуры нейросетевого классификатора. В качестве целевой функции предложена скалярная свертка критериев по нелинейной схеме компромиссов

**Ключевые слова:** многокритериальная оптимизация, архитектура, нейронные сети, классификатор.

**ACM Classification Keywords:** H.1 Models and Principles – H.1.1 – Systems and Information Theory; H.4.2 – Types of Systems; C.1.3 Other Architecture Styles – Neural nets

#### Введение

Настоящая работа является развитием результатов, полученных в работе [1]. Как и в [1], будем рассматривать структуру  $q$ -слойного нейросетевого классификатора с прямыми связями, представленную на рис.1.

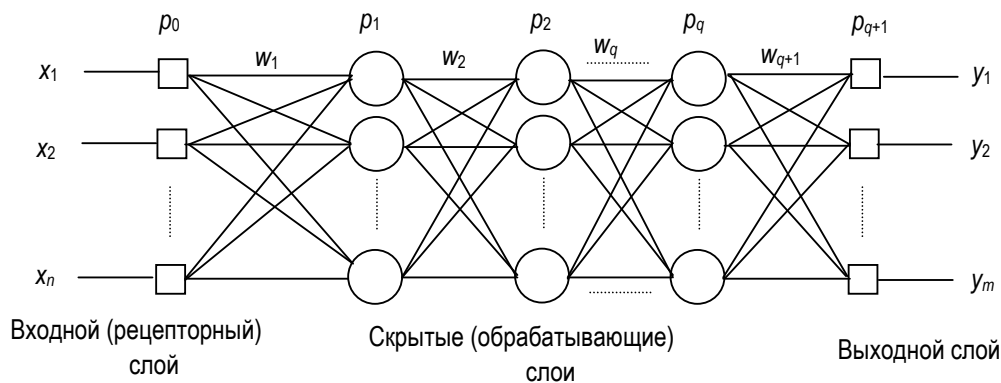


Рис.1

Здесь  $x_1, x_2, \dots, x_n$  – признаки объекта классификации, составляющие входной вектор  $x = \{x_i\}_{i=1}^n$ ;  $p_0 = n$  – число нейронных элементов в рецепторном слое;  $p_1, p_2, \dots, p_q$  – число нейронов в каждом из скрытых слоев;  $p_{q+1} = m$  – число нейронов в выходном слое (число классов);  $y = \{y_k\}_{k=1}^m$  – выходной вектор нейронной сети, определяющий принадлежность объекта

классификации одному из  $m$  классов;  $w_1, w_2, \dots, w_q, w_{q+1}$  – векторы синаптических весов нейронной сети.

Необходимые сведения из теории нейронных сетей содержатся в [2-4]. На рис.2 представлена структурная схема нейрона.

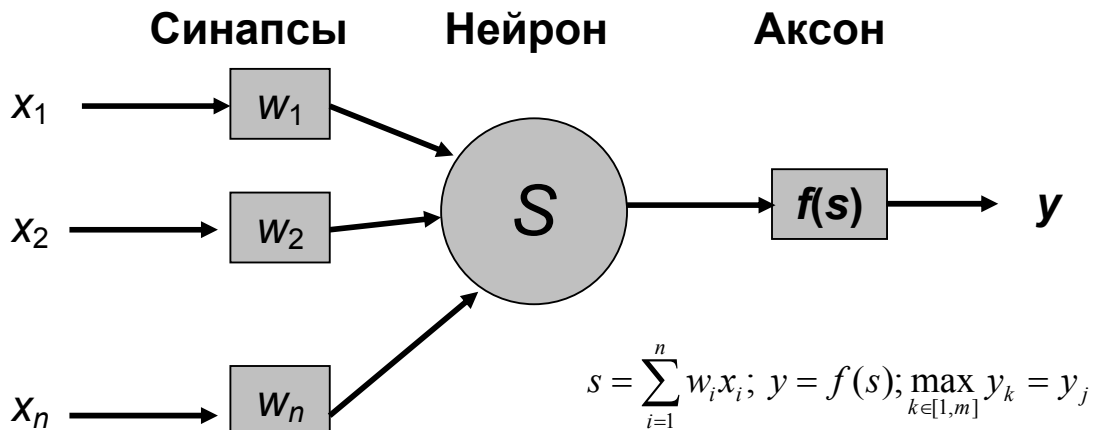


Рис. 2

Количество нейронов входного слоя  $p_0 = n$  определяется размерностью входного вектора признаков и не подлежит изменениям. Аналогично, количество нейронов выходного слоя  $p_{q+1} = m$  определяется числом областей (классов), на которые делится пространство признаков и тоже является постоянным. Количество же обрабатывающих (скрытых) слоев  $q$  и число нейронов в каждом из них  $p_1, p_2, \dots, p_q$  составляют понятие **архитектуры** [2] нейронной сети и могут служить аргументами (независимыми переменными) при ее оптимизации.

В настоящей работе ограничим исследование случаем, когда число  $q$  является фиксированным и заданным. Тогда аргументами оптимизации архитектуры нейронного классификатора являются количества нейронов в каждом из обрабатывающих слоев, составляющие вектор независимых переменных  $p = \{p_j\}_{j=1}^q$ . От выбора архитектуры  $p$  зависит качество функционирования нейронного классификатора.

Проблема заключается в таком выборе архитектуры, при котором нейронный классификатор в заданных условиях функционирования характеризуется наилучшими свойствами.

### Постановка задачи

В общем виде проблема может быть формально представлена задачей

$$p^* = \arg \operatorname{extr}_{p \in P} Y(p) \quad (1)$$

где  $Y(p)$  – целевая функция;  $\operatorname{extr}_{p \in P}$  – оператор экстремизации целевой функции по аргументам  $p$ ;  $P$

– допустимая область независимых переменных.

Для конструктивного решения задачи сделаем дополнительные частные предположения. Каждому свойству нейронного классификатора поставим в соответствие количественную характеристику  $f(p)$ , имеющую смысл критерия качества его функционирования. Одним из таких критериев является вероятность ошибки классификации. Будем определять этот критерий экспериментально и приближенно представим его как количество ошибок классификации  $e(p)$ , отнесенное к общему, достаточно большому количеству испытаний  $N$ :

$$f_1(p) = \frac{e(p)}{N} \quad (2)$$

Предполагается, что с ростом в некоторых разумных пределах числа нейронов в обрабатывающих слоях точность классификации повышается, и величина этого критерия уменьшается. Предельно допустимое значение ошибки сети должно быть известно из физических соображений и задано как ограничение

$$f_1(p) \leq A_1.$$

Второй критерий характеризует время, необходимое для обучения нейронной сети при данной архитектуре  $p$ . Наблюдается тесная корреляция между таким временем и суммарным количеством нейронов в скрытых слоях классификатора. Поэтому представим этот критерий в виде

$$f_2(p) = \sum_{k=1}^q p_k \quad (3)$$

Отметим, что данным критерием характеризуется и время прохождения сигнала через нейронную сеть от входа к выходу. С ростом числа нейронов значение критерия увеличивается. Предельно допустимое значение второго критерия определяется допустимым временем обучения нейронной сети и задается как ограничение  $f_2(p) \leq A_2$ .

Существуют и другие критерии для характеристики различных свойств нейронного классификатора. В данной работе мы ограничимся только приведенными двумя основными критериями, имея в виду, что излагаемая методика допускает включение в рассмотрение и других свойств классификатора.

Допустимая область аргументов оптимизации задается параллелепипедным ограничением  $P = \{p | 0 \leq p_k \leq P_u, k \in [1, P_u], u \in [1, q]\}$ , где  $P_u$  – максимальное число нейронов в  $u$ -м слое.

Поскольку оба включенных в рассмотрение критерия подлежат минимизации (чем критерий меньше, тем лучше соответствующее свойство классификатора), то оператор экстремизации целевой функции приобретает вид:  $extr = \min_{p \in P}$ .

Таким образом, оба критерия являются противоречивыми, неотрицательными, минимизируемыми и ограниченными. Налицо все предпосылки для использования в качестве целевой функции скалярной свертки критериев по нелинейной схеме компромиссов [4]. Такая свертка в унифицированной версии выражается формулой

$$Y(p) = Y[f(p)] = \frac{A_1}{A_1 - f_1(p)} + \frac{A_2}{A_2 - f_2(p)} \quad (4)$$

где  $f(p) = \{f_r(p)\}_{r=1}^{r=2}$  – двумерный вектор частных критериев. Учитывая (2), (3) и (4), выражение (1) для задачи оптимизации архитектуры нейронного классификатора преобразуется к виду

$$p^* = \arg \min_{p \in P} \left[ \frac{A_1}{A_1 - e(p)/N} + \frac{A_2}{A_2 - \sum_{k=1}^q p_k} \right] \quad (5)$$

Нетрудно видеть, что в формуле (5) зависимость  $e(p)$  априори является неизвестной и подлежащей экспериментальному определению.

### Методы оптимизации целевой функции

В работе [1] показано, что задача (5) является задачей нелинейного дискретного (целочисленного) программирования. Указаны предположения, при которых эта задача может решаться хорошо разработанными непрерывными методами.

В том случае, когда компоненты возможных решений многокритериальных задач могут принимать только дискретные значения  $p_k^{(P_u)}$ ,  $k \in [1, P_u]$ ,  $u \in [1, q]$ , скалярная свертка по нелинейной схеме компромиссов  $Y(p)$  является *решетчатой* функцией, заданной на дискретном множестве  $P$ . Оптимизация решетчатой целевой функции, построенной по нелинейной схеме компромиссов, сводится к задаче нелинейного программирования с дискретными (целочисленными) аргументами, решение которой, как отмечено выше, достаточно сложно.

Для решения этой проблемы мы предполагаем, что при дискретном множестве  $P$  существует вспомогательная область непрерывных аргументов  $p_c \in P_c$ , которая содержит все дискретные точки  $p_k^{(P_u)}$  и всё непрерывное пространство между ними. В области  $P_c$  определена непрерывная функция  $Y(p_c)$ , которая в точках  $p_k^{(P_u)}$  совпадает с решетчатой функцией  $Y(p)$ .

Это предположение позволяет получить аналитическое решение, если в выражении (5) зависимость  $e(p)$  задана, например, в виде регрессионной модели. Тогда можно воспользоваться необходимым условием минимума функции:  $\frac{\partial Y(p_c)}{\partial p_c} = 0$ . Решение этой системы уравнений дает компромиссно-

оптимальную непрерывную точку  $p_c^*$ . Последний шаг алгоритма – поиск на  $P$  ближайшей к  $p_c^*$  возможной *дискретной* точки, которая и будет искомым дискретным решением  $p^*$ . В нашем случае, к сожалению, задание аналитических зависимостей весьма затруднительно или вообще невозможно.

Мы рассматриваем как основной случай, когда функции  $e(p)$  и, следовательно  $Y(p)$ , неизвестны, но есть возможность определять значения функции  $Y(p)$  в точках  $p_k^{(P_u)}$  измерением или вычислением. Тогда можно организовать натурный или вычислительный эксперимент, в результате которого осуществляется поисковое движение к искомой дискретной компромиссно-оптимальной точке  $p^*$ .

Возможны различные подходы к организации поисковой процедуры, которая должна давать последовательность улучшающихся решений. Один из них – дискретный аналог метода симплекс-планирования Нелдера-Мида (метод деформируемого многогранника) [5]. Это – разновидность градиентных методов, весьма часто и успешно применяющихся на практике.

Проведем исследование для  $q=2$ . На плоскости аргументов  $p_1 - p_2$  в некоторой стартовой области строится исходный регулярный симплекс, который в двумерном случае представляет собой равнобедренный треугольник с вершинами  $p^{(1)}, p^{(2)}, p^{(3)}$ . Для архитектур стартового симплекса вычисляются значения скалярных сверток  $Y^{(1)}, Y^{(2)}, Y^{(3)}$ . Сравнивая между собой эти значения, определяются худшее. С большой вероятностью можно утверждать, что архитектура, полученная зеркальным отображением худшей в исходном симплексе точки относительно центра противоположной грани, окажется лучше. Исходя из этого, образуется новый симплекс. Процедура построения симплексов продолжается до тех пор, пока мы получим архитектуру  $p^*$ , соответствующую минимуму целевой функции [1].

Рассмотренный метод предусматривает старт поисковой процедуры от архитектуры, которая, по мнению разработчика, находится достаточно близко к оптимальной точке. Если в процессе поиска имеет место возрастание числа нейронов в обрабатываемых слоях, то теория нейронных сетей [2] характеризует данный подход как *конструктивный*. При избыточном стартовом количестве нейронов подход именуется *деструктивным* (принцип Родена: чтобы изваять скульптуру, нужно взять целую глыбу мрамора и удалить из нее лишнее).

Осуществление изложенных в работе этапов векторной оптимизации позволяет получить архитектуру нейросетевого классификатора, при которой системно увязываются противоречивые критерии эффективности его функционирования.

---

### **Обучение нейросетевого классификатора текстов**

---

В качестве примера рассмотрим в общих чертах задачу многокритериальной оптимизации архитектуры нейросетевого классификатора текстов. Система текстовой классификации [4] состоит из двух основных частей: частотный анализатор с системным словарем и собственно нейросетевой классификатор. Структурная схема, а также этапы работы нейросетевого классификатора приведены в [1].

Обучение нейросетевого классификатора текстов было реализовано при помощи программы, написанной на языке программирования Delphi. (Рис. 4)

В данной программе использовался метод обучения с обратным распространением ошибки. Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы.

На основе имеющегося множества учебных текстов (легальных писем и спама) создается системный словарь, который содержит в себе множество слов, встречающихся во всех учебных текстах, а также их частотную характеристику. При помощи кнопки «Загрузка матриц» формируется файл с весовыми коэффициентами нейронной сети, после чего можно приступить непосредственно к этапу обучения, в процессе которого весовые коэффициенты будут регулироваться программой, таким образом настраивая нейронную сеть для дальнейшей работы. Настроенную нейронную сеть можно использовать для классификации различных текстов. Также данная программа позволяет регулировать размерность нейронной сети и другие аргументы, использующиеся при вычислениях.

---

### **Многокритериальная оптимизация нейросетевого классификатора текстов**

---

Для оптимизации архитектуры нейросетевого классификатора использовалось специальное программное обеспечение (Рис. 5)

В данной программе реализована возможность проведения оптимизации с использованием разных методов и в разных режимах. Режим «Эксперимент» позволяет провести оптимизацию в том случае, когда зависимости критериев от параметров не известны и для определения значения критериев необходим эксперимент. Режим «Аналитика» используется в том случае, когда критерии выражаются через

параметры явно, то есть, когда известны функциональные зависимости критериев от параметров. В данной работе будут использованы оба режима.

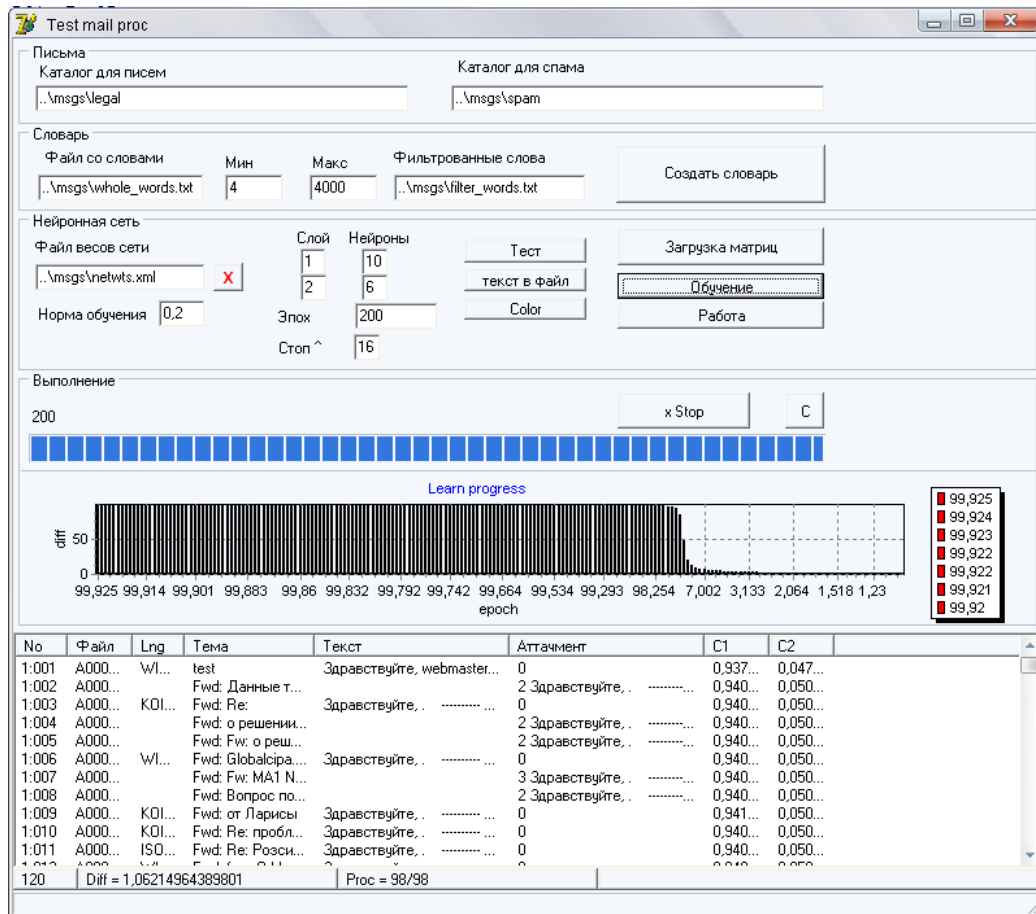


Рис. 4

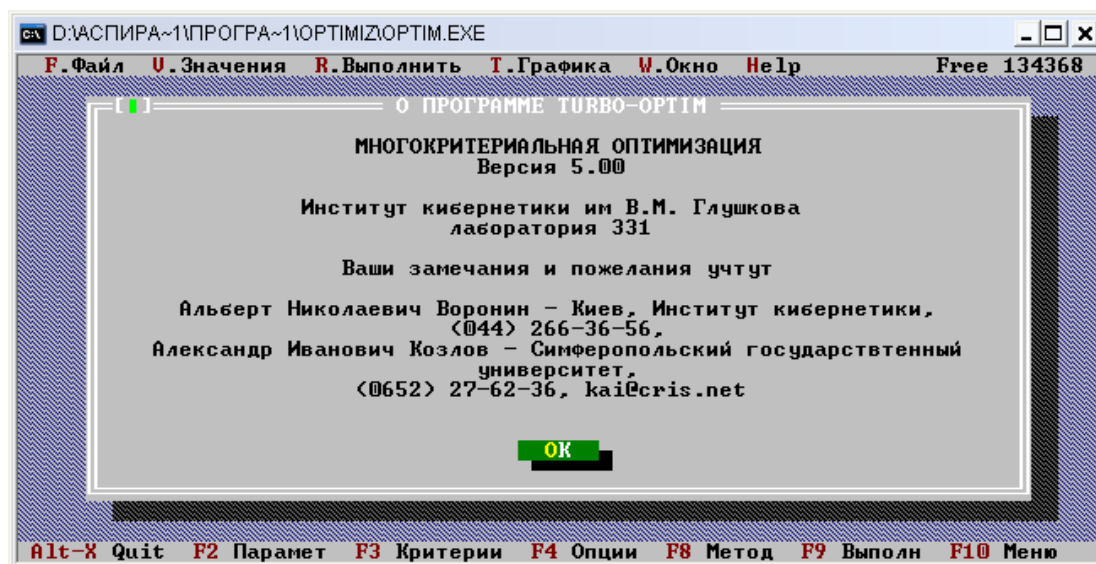


Рис. 5

Как уже упоминалось выше, в качестве критериев оптимизации были выбраны:

- Время обучения нейронной сети, пропорциональное суммарному количеству нейронов в сети, аналитическое выражение критерия выглядит следующим образом:  $f_2(p) = \sum_{k=1}^q p_k$ .
- Вероятность возникновения ошибки  $f_1(p)$ . Аналитическое выражение на начальном этапе проведения процедуры оптимизации является неизвестным.

Поскольку аналитические выражения известны не для всех критериев, на начальной стадии проводится оптимизация архитектуры нейросетевого классификатора в режиме «Эксперимент». В данном режиме процедура оптимизации проводится пошагово и требует знание значений критерия, определяющего вероятность возникновения ошибки, для каждой итерации. Значения данного критерия в каждом шаге итерации вычисляются в программе обучения нейросетевого классификатора.

Как уже упоминалось ранее, для проведения поисковой процедуры нахождения оптимальной архитектуры нейросетевого классификатора был выбран поисковый метод симплекс-планирования.

Используя программу обучения нейронной сети и программу оптимизации для метода симплекс-планирования, получим следующую таблицу значений:

Таблица 1

<i>и</i> - номер испытания	<i>p</i> <sub>1</sub> - количество нейронов в 1-ом слое	<i>p</i> <sub>2</sub> - количество нейронов в 2-ом слое	<i>e</i> (%) - значение ошибки классификации	<i>t</i> (с) - время обучения, $f_2(p) = \sum_{k=1}^q p_k$
1	6	6	99.9	12
2	9	7	3.87	16
3	7	9	99.0	16
4	10	10	0.376	20
5	13	8	0.274	21
6	11	8	0.533	19
7	11	5	1.6	16
8	9	3	99.8	12
9	11	7	0.67	18
10	9	5	99.6	14
11	10	6	1.19	16
12	11	4	99.7	15
13	10	6	1.19	16
14	11	5	1.6	16
15	11	7	0.67	18
16	11	5	1.6	16
17	11	6	0.95	17
18	11	5	1.6	16
19	11	6	0.95	17
20	11	6	0.95	17

Для аналитического выражения критерия, определяющего вероятность возникновения ошибки, была выбрана содержательная модель, имеющая вид:  $f_1(p_1, p_2) = a_1 / p_1 + a_2 / p_2$ , где  $a_1, a_2$  – коэффициенты регрессии,  $p_1, p_2$  – количество нейронов соответственно в 1-ом и 2-ом слоях.

Для нахождения коэффициентов регрессии составляется функция суммы квадратов невязок:



$$E(a) = \sum_{u=1}^N \left( \frac{a_1}{p_1^{(u)}} + \frac{a_2}{p_2^{(u)}} - f_1^{(u)} \right)^2,$$

минимизация которой, в соответствии с методом наименьших квадратов, дает неизвестные коэффициенты регрессии.

Таким образом, для метода симплекс-планирования было получено аналитическое выражение критерия возникновения ошибки:

$$y_1 = \frac{9.97}{p_1} + \frac{3.7}{p_2}.$$

Подставив аналитические выражения для критериев в режиме «аналитика», получим следующие значения аргументов, определяющие компромиссно-оптимальную архитектуру нейросетевого классификатора, найденные методом симплекс-планирования:

$p_1$  – количество нейронов в 1-ом слое,  $p_1^* = 9$ ;

$p_2$  – количество нейронов во 2-ом слое,  $p_2^* = 6$ .

Таким образом, видно, что полученный результат подтверждает экспериментальные данные, представленные в Таблице 1. То есть использование метода симплекс-планирования для нахождения оптимальной архитектуры нейросетевого классификатора дает компромиссно-оптимальную модель, при которой системно увязываются противоречивые критерии эффективности его функционирования: время обучения и вероятность возникновения ошибки.

---

### Благодарности

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

---

### Библиография

1. Альберт Воронин, Юрий Зиатдинов, Анна Антонюк, Многокритериальная оптимизация архитектуры нейросетевых классификаторов // International Book Series Number 8, P.32-39, ITHEA Sofia, 2009.
2. Бодянский Е.В., Руденко О.Г. Искусственные нейронные сети: архитектуры, обучение и применение. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с.
3. Головкин В.А. Нейронные сети: обучение, организация и применение. – М.: ИПРЖР, 2001. – 256 с.
4. Борисов В.С. Самообучающийся классификатор текстов на естественном языке // Кибернетика и системный анализ – 2007. – №3. – С.169-176.
5. Воронин А.Н., Зиатдинов Ю.К., Козлов А.И. Векторная оптимизация динамических систем. – Киев: Техніка, 1999. – 284 с.
6. Воронин А.Н., Мосорин П.Д., Ясинский А.Г. Многокритериальные задачи оптимизации с дискретными аргументами // Автоматика-2000. Міжнародна конференція з автоматичного управління: Праці. –Т.1. – Львів: ДНДІІ, 2000. – С. 75-78.

---

### Сведения об авторах

**Воронин Альберт Николаевич** – профессор, доктор технических наук, профессор кафедры компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: [alnv@voliacable.com](mailto:alnv@voliacable.com)

**Зиатдинов Юрий Кашафович** – профессор, доктор технических наук, заведующий кафедрой компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: [oberst@nau.edu.ua](mailto:oberst@nau.edu.ua)

**Антонюк Анна Александровна** – аспирант Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: [niuriel@mail.ru](mailto:niuriel@mail.ru)

## КОРРЕКТНЫЕ МОДЕЛИ АЛГОРИТМОВ РАСПОЗНАВАНИЯ

Виктор Краснопрошин, Владимир Образцов

**Реферат:** Рассматривается задача распознавания образов с обучением. Определяются общие условия существования корректных алгоритмов, и показывается, что их построение сводится к решению систем линейных матричных неравенств.

**Ключевые слова:** Задача распознавания образов с обучением, корректные модели алгоритмов распознавания, условия корректности моделей распознавания.

**Ключевые слова АСМ-классификации:** Categories: 1. Computing Methodologies; 1.5 PATTERN RECOGNITION; 1.5.1 Models; Subject descriptor: Models Deterministic

### Введение

Понятие корректных алгоритмов в теории распознавании образов было введено в конце 80-ых годов в работе [Журавлев, 1978]. Результаты их исследований подвели, в некотором смысле, черту в вопросах построения оптимальных алгоритмов для решения задач распознавания с обучением. Фактически было показано, что оптимальность (корректность) алгоритмов на любой конечной выборке носит локальный характер и достижима в алгебраических расширениях широкого класса эвристических алгоритмов.

Связь локальных и глобальных характеристик задачи распознавания на языке теории категорий исследовалась в [Рудаков, 1989]. Аналогичные зависимости, но уже на языке теории множеств, установлены в [Краснопрошин, 2006]. Было показано, что эта связь не может быть конструктивно реализованной, так как при любых условиях природа задачи распознавания остается индуктивной.

Что же касается проблем построения корректных алгоритмов, то здесь по-прежнему остаются некоторые вопросы. В частности, с практических позиций важно знать каковы границы сложности построения таких алгоритмов. В данной работе показывается, что в общем случае задача построения корректных алгоритмов сводится к решению систем линейных матричных неравенств. И, следовательно, нижняя граница сложности равна  $O(n^3)$  \*).

### Постановка Задачи. Основные Понятия и Определения

Рассмотрим задачу распознавания образов с обучением [Журавлев, 1978]. Пусть в некотором пространстве задано произвольное множество допустимых объектов  $\{S\}$ . Предположим, что  $\{S\}$  покрывается конечным числом подмножеств (к л а с с о в )  $K_1, \dots, K_l$  (т.е.  $\{S\} \subset \bigcup_{i=1}^l K_i$ ) таких, что  $K_i \neq \emptyset$  для всех  $i = 1, 2, \dots, l$ . При этом подмножества  $K_i$  определены не полностью – имеется лишь конечное число допустимых объектов  $\{S_0\}$  известной классификации.

На множестве  $\{S\}$  введем систему предикатов  $P_j(S) (j = 1, 2, \dots, l)$  такую, что

$$P_j(S) = \begin{cases} 1, & \text{если } S \in K_j, \\ 0, & \text{иначе.} \end{cases}$$

Тогда информацию о принадлежности произвольного объекта  $S \in \{S\}$  к классам  $K_1, \dots, K_l$  можно представить в виде вектора  $\alpha(S) = (\alpha_1(S), \dots, \alpha_l(S))$ , где  $\alpha_j(S) = P_j(S)$ . Согласно предположениям, для объектов  $S \in \{S_0\}$  векторы  $\alpha(S)$  считаются заданными и называются и н ф о р м а ц и о н н ы м и .

\*)  $n$  – число объектов контрольной (тестовой) выборки.

В множестве  $\{S_0\}$  выделим некоторую совокупность допустимых объектов  $\tilde{S}_m = \{S_1, \dots, S_m\}$  такую, что

$$(\tilde{S}_m \cap K_i \neq \emptyset) \forall i \in \{1, 2, \dots, l\}.$$

Упорядоченный набор  $(S_1, \alpha(S_1), \dots, S_m, \alpha(S_m))$  образует так называемую стандартную начальную информацию о классах  $K_1, \dots, K_l$ , обозначим ее через  $I_0(K_1, \dots, K_l)$  (или  $I_0$ )\*.

С учетом введенных обозначений основную задачу распознавания образов с обучением можно сформулировать следующим образом. Пусть задана некоторая информация  $I_0$  и произвольный набор допустимых объектов  $\tilde{S}^q = \{S^1, \dots, S^q\}$ . Требуется указать алгоритм (метод, правило), который на основании информации  $I_0$  для любого объекта  $S^i \in \tilde{S}^q$  ( $i = 1, 2, \dots, q$ ) определяет принадлежность последнего к классам  $K_1, \dots, K_l$ . Обозначим эту задачу через  $Z$ , и в дальнейшем будем отождествлять ее с парой  $(I_0, \tilde{S}^q)$ , т.е.  $Z = (I_0, \tilde{S}^q)$ .

Алгоритм  $A$ , решающий сформулированную задачу, называется распознающим. Результатом работы такого алгоритма при распознавании произвольного допустимого объекта  $S$  является вектор

$$\alpha^A(S) = (\alpha_1^A(S), \dots, \alpha_l^A(S)), \text{ где } \alpha_j^A(S) \in \{1, 2, \dots, l\},$$

или матрица

$$\|\alpha_j^A(S^i)\|_{q \times l},$$

в случае, если классифицируется совокупность объектов  $\tilde{S}^q = \{S^1, \dots, S^q\}$ . Вектор  $\alpha_j^A(S)$  называется классификационным, его значения интерпретируются следующим образом

$$\alpha_j^A(S) = \begin{cases} 0, & \text{если } A \text{ не заносит } S \text{ в класс } K_j, \\ 1, & \text{если } A \text{ заносит } S \text{ в класс } K_j, \\ 2, & \text{если } A \text{ отказался от вычисления } P_j(S). \end{cases}$$

Понятно, что алгоритм  $A$  можно рассматривать как отображение, определенное на множестве  $I_0 \times \{S\}$ , т.е.

$$\forall S \in \{S\} (A: I_0 \times S \rightarrow (\alpha_1^A(S), \dots, \alpha_l^A(S))). \quad (1)$$

Разработано значительное число подходов к решению сформулированной задачи. Рассмотрим один из них, предложенный в [Журавлев, 1978].

Пусть задано произвольное семейство (модель) алгоритмов распознавания  $\{A\}$  и некоторая стандартная начальная информация  $I_0$ . Предположим, что в множестве  $\{S_0\}$  некоторым образом выделена конечная совокупность объектов  $\tilde{S}^q$  ( $q \in \mathbb{N}$ ), для которых имеет место

$$\begin{cases} (\tilde{S}_m \cap \tilde{S}^q = \emptyset), \\ (\tilde{S}^q \cap K_i \neq \emptyset) \forall i \in \{1, 2, \dots, l\}. \end{cases}$$

Такая выборка называется контрольной. Матрицу ее информационных векторов обозначим  $I(\tilde{S}^q)$ .

Суть рассматриваемого подхода заключается в построении алгоритмов (1), точных на  $Z = (I_0, \tilde{S}^q)$  в

\* ) Отметим, что существуют и другие способы задания начальной информации о классах. В данном случае выборка  $\tilde{S}_m$  обычно называется обучающей.

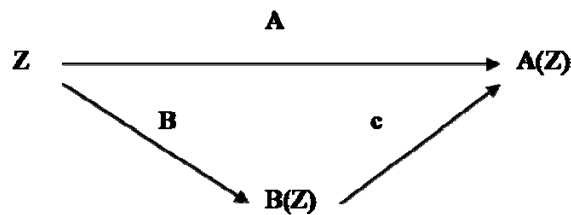
смысле следующего определения. Алгоритм  $A \in \{A\}$  называется корректным на  $Z$ , если  $A(Z) = I(\tilde{S}^q)$ . Модель  $\{A\}$ , которая содержит такой алгоритм, также называется корректной.

Понятно, что в этом подходе основной проблемой является определение условий на задачу  $Z$  и модель алгоритмов  $\{A\}$ , при которых последняя корректна на  $Z$ . Прежде, чем перейти к ее решению, уточним множества рассматриваемых в дальнейшем алгоритмов и задач.

Пусть  $\mathbb{R}^q$ ,  $\mathbb{B}_3^q$  - пространства матриц размерности  $q \times l$  над полем вещественных чисел  $\mathbb{R}$  и множеством  $\mathbb{B}_3 = \{0,1,2\}$  соответственно. В [Журавлев, 1978] показано, что любой алгоритм  $A \in \{A\}$  вида (1) представим на  $Z$  в виде суперпозиции

$$A(Z) = (c \circ B)(Z),$$

т.е. коммутативна диаграмма



где  $B(Z) \in \mathbb{R}^q$ ,  $A(Z) \in \mathbb{B}_3^q$ . Подалгоритмы

$$B \in \mathfrak{M} \quad (B : (I_0, \tilde{S}^q) \rightarrow \mathbb{R}^q), \quad c \in C \quad (c : \mathbb{R}^q \rightarrow \mathbb{B}_3^q),$$

называются соответственно распознающими операторами ( $\mathfrak{M}$ ) и решающими правилами ( $C$ ). Такое представление позволяет при рассмотрении алгоритмов ограничиться множеством

$$A_{\mathfrak{M}} = c \circ \mathfrak{M},$$

порожденным суперпозицией произвольных распознающих операторов  $\mathfrak{M}$  и решающих правил  $C$ .

Пусть  $\mathbb{B}_2^q$  - пространство матриц размерности  $q \times l$  над  $\mathbb{B}_2 = \{0,1\}$ . Нетрудно заметить, что при фиксированных начальной информации  $I_0$  и конечном размере  $q$  контрольной выборки  $\tilde{S}^q$  (величина  $q < +\infty$  принципиального значения не имеет), множество задач  $Z = (I_0, \tilde{S}^q)$  можно однозначно определить по матрице  $I(\tilde{S}^q)$ . В соответствии с этим введем множество

$$Z_2^q = \{(I_0, \tilde{S}^q) \mid \bigcup_{\tilde{S}^q} I(\tilde{S}^q) = \mathbb{B}_2^q\}$$

и рассмотрим условия корректности произвольной модели  $A_{\mathfrak{M}}$  на множестве задач  $Z_2^q$ .

Пусть  $\mathfrak{M}$  - некоторая совокупность (м о д е л ь) распознающих операторов и  $Z$  - произвольная задача из  $Z_2^q$ . Очевидно, что решения  $\mathfrak{M}(Z)$  задачи  $Z$  операторами  $B \in \mathfrak{M}$  образует подмножество в пространстве  $\mathbb{R}^q$ , т.е.  $\mathfrak{M}(Z) \subseteq \mathbb{R}^q$ . Зафиксируем решающее правило  $c \in C$ . По матрице информационных векторов  $I(\tilde{S}^q)$ , соответствующих контрольной выборке  $\tilde{S}^q$  задачи  $Z$ , в пространстве  $\mathbb{R}^q$  построим множество

$$R_c(Z) = \{R \mid R \in \mathbb{R}^q, c(R) = I(\tilde{S}^q)\},$$

и назовем его областью корректности решающего правила  $c \in C$  на задаче  $Z$ .

С учетом проведенных построений получаем, что модель  $A_{\mathfrak{M}}$  корректна на  $Z \in Z_2^{q_l}$  в том и только том случае, если

$$\mathfrak{M}(Z) \cap R_c(Z) \neq \emptyset. \quad (2)$$

Нетрудно показать, что данное условие позволяет фактически исключить из рассмотрения решающее правило  $c \in C$ . Действительно, из него с необходимостью следует

$$R_c(Z) \neq \emptyset.$$

Но последнее справедливо на всем множестве  $Z_2^{q_l}$  в том и только том случае, если для решающего правила  $c \in C$  имеет место

$$\forall b \in \mathbb{B}_2^{q_l} \exists R \in \mathbb{R}^{q_l} (c(R) = b). \quad (3)$$

Решающее правило  $c \in C$ , удовлетворяющее условию (3), называется корректным. Например, линейные пороговые решающие правила, определенные в пространстве  $\mathbb{R}^{q_l}$

$$\forall R = \|r_{ij}\| \in \mathbb{R}^{q_l} (c(R) = \|c(r_{ij})\| \in \mathbb{B}_2^{q_l}), \text{ где } (i, j) \in \{1, 2, \dots, q\} \times \{1, 2, \dots, l\}, \text{ и}$$

$$c(r_{ij}) = \begin{cases} 0, & \text{если } r_{ij} < c_0, \\ 1, & \text{если } r_{ij} > c_1, \\ 2, & \text{если } c_0 \leq r_{ij} \leq c_1, \end{cases}$$

является корректным при условии  $c_0 \leq c_1$  (где  $c_0, c_1 \in \mathbb{R}$ ). Множество таких решающих правил обозначим  $C(c_0, c_1)$ . Легко убедиться, что при исследовании вопросов корректности можно ограничиться моделями алгоритмов, которые порождаются операторами из некоторой модели  $\mathfrak{M}$  и решающими правилами из множества  $C(c_0, c_1)$ .

Таким образом, проведенные рассуждения позволяют сформулировать следующую задачу. Требуется определить условия на задачу  $Z \in Z_2^{q_l}$  и модель распознающих операторов  $\mathfrak{M}$ , при которых последняя удовлетворяет условию (2) для некоторого решающего правила  $c \in C(c_0, c_1)$ .

### Общие Условия Корректности Моделей Алгоритмов Распознавания

Пусть  $R_c(Z)$  - область корректности фиксированного решающего правила  $c \in C(c_0, c_1)$  на задаче  $Z = (I_0, \tilde{S}^q)$ . Покажем, что  $R_c(Z)$  представляет собой множество решений системы строгих линейных неравенств.

В соответствии с матрицей информационных векторов  $I(\tilde{S}^q) = \|\alpha_{ij}\|_{q \times l}$  контрольной выборки  $\tilde{S}^q$ , множество индексов  $I = \{1, 2, \dots, q\} \times \{1, 2, \dots, l\}$  разобьем на подмножества

$$M_t = \{(i, j) \mid (i, j) \in I, \alpha_{ij} = t\}.$$

Т.к.  $I(\tilde{S}^q) \in \mathbb{B}_2^{q_l}$ , то  $t \in \{0, 1\}$  и для введенных подмножеств имеют место соотношения

$$\begin{cases} M_0 \cup M_1 = I, \\ M_0 \cap M_1 = \emptyset. \end{cases}$$

Покажем, что  $R_c(Z)$  образует выпуклое подмножество в пространстве  $\mathbb{R}^{q_l}$ , то есть

$$\forall R_k = \left\| r_{ij}^k \right\| \in R_c(Z) (k=1,2) \forall \lambda \in [0,1] (\lambda R_1 + (1-\lambda)R_2 \in R_c(Z)).$$

Действительно, пусть  $R_1$  и  $R_2$  - произвольные матрицы из  $R_c(Z)$ . Учитывая определения множеств  $C(c_0, c_1)$  и  $R_c(Z)$ , имеем

$$\forall (i, j) \in M_1 \forall \lambda \in [0,1] (\lambda r_{ij}^1 + (1-\lambda)r_{ij}^2 > \lambda c_1 + (1-\lambda)c_1 = c_1),$$

$$\forall (i, j) \in M_0 \forall \lambda \in [0,1] (\lambda r_{ij}^1 + (1-\lambda)r_{ij}^2 < \lambda c_0 + (1-\lambda)c_0 = c_0).$$

Объединяя последние неравенства, получаем

$$\forall \lambda \in [0,1] (\lambda R_1 + (1-\lambda)R_2 \in R_c(Z)).$$

В силу выбора матриц  $R_1$  и  $R_2$ , из данного включения следует выпуклость области корректности  $R_c(Z)$ . Но это, в свою очередь, означает, что  $R_c(Z)$  является решением некоторой системы неравенств.

Убедимся, что эта система линейна. По аналогии с векторным случаем, введем скалярное произведение матриц в пространстве  $\mathbb{R}^{qI}$  следующим образом

$$\forall R_1, R_2 \in \mathbb{R}^{qI} (\langle R_1, R_2 \rangle = \sum_{(i,j) \in I} r_{ij}^1 \cdot r_{ij}^2).$$

Рассмотрим множество

$$H(R_0, a) = \{X \in \mathbb{R}^{qI} \mid \langle R_0, X \rangle = a\},$$

где  $a \in \mathbb{R}$ ,  $R_0 \in \mathbb{R}^{qI}$ . Легко видеть, что при фиксированных  $a$  и ненулевой матрице  $R_0$ , оно представляет собой матричный аналог гиперплоскости векторного пространства. В свою очередь, множества

$$H^+(R_0, a) = \{X \in \mathbb{R}^{qI} \mid \langle R_0, X \rangle > a\}, \quad H^-(R_0, a) = \{X \in \mathbb{R}^{qI} \mid \langle R_0, X \rangle < a\},$$

образуют открытые полупространства, на которые гиперплоскость  $H(R_0, a)$  разделяет пространства  $\mathbb{R}^{qI}$ .

Обозначим  $\{E_{ij}\}_{(i,j) \in I}$  - канонический базис в  $\mathbb{R}^{qI}$  и рассмотрим множество

$$H_0 = \left( \bigcap_{(i,j) \in M_1} H^+(E_{ij}, c_1) \right) \cap \left( \bigcap_{(i,j) \in M_0} H^-(E_{ij}, c_0) \right),$$

где  $c_0, c_1$  - пороги решающего правила  $c \in C(c_0, c_1)$ . Нетрудно проверить, что  $H_0$  - выпуклое и непустое подмножество в пространстве  $\mathbb{R}^{qI}$ .

Выберем некоторую матрицу  $R \in R_c(Z)$ . Из определений множества  $C(c_0, c_1)$  и области корректности, имеем

$$\forall (i, j) \in M_1 (\langle E_{ij}, R \rangle > c_1) \Rightarrow R \in H^+(E_{ij}, c_1), \quad \forall (i, j) \in M_0 (\langle E_{ij}, R \rangle < c_0) \Rightarrow R \in H^-(E_{ij}, c_0).$$

Но это означает, что  $R \in H_0$ . Откуда, с учетом произвольности выбора матрицы  $R$  следует

$$R_c(Z) \subseteq H_0.$$

Нетрудно убедиться и в обратном. Таким образом,

$$R_c(Z) = H_0,$$

то есть  $R_c(Z)$  образовано пересечением  $q \cdot I$  открытых полупространств в  $\mathbb{R}^{qI}$ .

Из приведенных выше рассуждений и определения множества  $H_0$  получаем, что область корректности  $R_c(Z)$  при фиксированном  $c \in C(c_0, c_1)$  и  $Z \in Z_2^{ql}$  является решением следующей системы линейных неравенств

$$\begin{cases} \langle E_{ij}, X \rangle > c_1, \forall (i, j) \in M_1, \\ \langle E_{ij}, X \rangle < c_0, \forall (i, j) \in M_0, \end{cases} \quad (4)$$

так как последняя однозначно определяет множество  $H_0$ . Отметим, что данная система единственна с точностью до изоморфизма базисов в пространстве  $\mathbb{R}^{ql}$ . В результате нами доказана следующая Л е м м а . Пусть  $\mathfrak{M}$  - произвольная модель распознающих операторов и  $c$  - решающее правило из  $C(c_0, c_1)$ . Модель  $\mathfrak{M}$  удовлетворяет условию (2) на  $Z \in Z_2^{ql}$  в том и только том случае, если существует такой оператор  $B \in \mathfrak{M}$ , что матрица  $B(Z)$  является решением системы (4).

Получим условия, при которых матрица  $B(Z) \in \mathbb{R}^{ql}$  является решением системы строгих линейных неравенств (4). Перейдем к системе, в которой все неравенства имеют одинаковый смысл и нулевую правую часть.

По матрице информационных векторов  $l(\tilde{S}^q)$ , соответствующей задаче  $Z \in Z_2^{ql}$ , в пространстве  $\mathbb{R}^{ql}$  введем одноместную операцию  $\alpha$  такую, что

$$\forall R = \|r_{ij}\| \in \mathbb{R}^{ql} \quad (R^\alpha = \|r_{ij}^\alpha\|),$$

где

$$r_{ij}^\alpha = \begin{cases} r_{ij}, & \text{если } \alpha_{ij} = 1, \\ -r_{ij}, & \text{если } \alpha_{ij} = 0. \end{cases}$$

Определим также в пространстве  $\mathbb{R}^{ql}$  матрицу  $R_c$  с элементами

$$r_{ij} = \begin{cases} c_1, & \text{если } \alpha_{ij} = 1, \\ c_0, & \text{если } \alpha_{ij} = 0. \end{cases}$$

Тогда, очевидно, систему неравенств (4) можно записать в виде

$$\langle E_{ij}, X^\alpha - R_c^\alpha \rangle > 0, \forall (i, j) \in I.$$

Т е о р е м а 1. Пусть  $\mathfrak{M}$  - произвольная модель распознающих операторов и  $c$  - решающее правило из  $C(c_0, c_1)$  с порогами  $0 \leq c_0 \leq c_1$ . Тогда модель  $\mathfrak{M}$  удовлетворяет условию (2) на  $Z \in Z_2^{ql}$  в том и только том случае, если

$$\exists B \in \mathfrak{M} \quad \forall R \in \mathbb{R}^{ql} \quad (R \geq 0, R \neq 0) \quad (\langle R, R_c^\alpha \rangle = 0 \Rightarrow \langle R, B^\alpha(Z) \rangle > 0). \quad (5)$$

Отметим, что условие (5) является конструктивным, так как может быть непосредственно использовано при построении корректных алгоритмов. Для разработки соответствующих методов необходимо конкретизировать вид модели распознающих операторов  $\mathfrak{M}$  и воспользоваться аппаратом теории систем конечных неравенств.

Множество

$$R' = \{R \in \mathbb{R}^{ql} \mid (R \geq 0, R \neq 0) \& (\langle R, R_c^\alpha \rangle = 0)\}$$

можно определить аналитически, используя свойства скалярного произведения и вид матрицы  $R_c$ . Действительно, рассмотрим

$$R^0 = \{R \in \mathbb{R}^{qf} \mid (r_{ij} \geq 0, \sum_{(i,j) \in I} r_{ij} = 1) \& ((\sum_{(u,v) \in M_0} r_{uv}) \cdot (\sum_{(i,j) \in M_1} r_{ij})^{-1} = (c_1) \cdot (c_0)^{-1})\}.$$

По построению последнего множества получаем

$$\forall R \in R^0 \quad (\langle R, R_c^\alpha \rangle = 0).$$

Нетрудно также заметить, что  $R^0$  взаимно-однозначно (с точностью до положительной числовой константы) определяет все множество  $R'$ . Это позволяет существенно упростить проверку условия (5).

**С л е д с т в и е .** В условиях теоремы 1 модель  $\mathfrak{M}$  удовлетворяет (2) на  $Z \in Z_2^{qf}$  в том и только том случае, если

$$\exists B \in \mathfrak{M} \quad \forall R \in R^0 \quad (\langle R, B^\alpha(Z) \rangle > 0).$$

Теорема 1 и следствие из нее справедливы для решающих правил  $c \in C(c_0, c_1)$  с порогами  $0 \leq c_0 \leq c_1$ . В то же время, используя отделимость выпуклого множества  $R_c(Z)$  в пространстве  $\mathbb{R}^{qf}$ , можно получить иные условия, при которых модель  $\mathfrak{M}$  удовлетворяет (2) на  $Z \in Z_2^{qf}$ .

Из построения множества  $H_0$  следует, что гиперплоскость  $H(E_{ij}, c_1)$  при  $(i, j) \in M_1$  отделяет полупространства  $H^-(E_{ij}, c_1)$  и  $H^+(E_{ij}, c_1)$ , а гиперплоскость  $H(E_{ij}, c_0)$  при  $(i, j) \in M_0$  отделяет полупространства  $H^-(E_{ij}, c_0)$  и  $H^+(E_{ij}, c_0)$ . Но это означает, что подмножество матриц  $R_c(Z) = H_0$  и только оно удовлетворяет в  $\mathbb{R}^{qf}$  условию

$$\min\{\langle R, E_{ij} \rangle \mid (i, j) \in M_1\} > c_1 \geq c_0 > \max\{\langle R, E_{ij} \rangle \mid (i, j) \in M_0\},$$

где  $R$  - произвольная матрица в  $\mathbb{R}^{qf}$ .

**Т е о р е м а 2.** Пусть  $\mathfrak{M}$  - некоторая модель распознающих операторов и  $c$  - решающее правило из  $C(c_0, c_1)$ . Тогда для  $\mathfrak{M}$  имеет место (2) на  $Z \in Z_2^{qf}$  в том и только том случае, если существует оператор  $B \in \mathfrak{M}$  такой, что матрица  $B(Z) = \|b_{ij}\|$  удовлетворяет одному из следующих условий:

$$\min_{(i,j) \in M_1} b_{ij} > c_1 \geq c_0 > \max_{(u,v) \in M_0} b_{uv},$$

$$\forall (i, j) \in M_1 \quad \forall (u, v) \in M_0 \quad (c_0 > 0) \quad (b_{ij} \cdot (c_1)^{-1} > b_{uv} \cdot (c_0)^{-1}),$$

$$\forall (i, j) \in M_1 \quad \forall (u, v) \in M_0 \quad (c_0 \neq c_1) \quad ((b_{ij} - c_0) \cdot (c_1 - c_0)^{-1} > 1) \& ((c_1 - b_{uv}) \cdot (c_1 - c_0)^{-1} > 1).$$

---

## Заключение

В работе получены общие условия корректности моделей алгоритмов распознавания  $A_{\mathfrak{M}}$  на множестве задач  $Z_2^{qf}$ . Нетрудно видеть, что эти условия позволяют свести проблему построения корректного алгоритма в  $A_{\mathfrak{M}}$  к решению систем неравенств в матричном пространстве  $\mathbb{R}^{qf}$ . Решение таких систем даже в случае линейной модели  $\mathfrak{M}$  является нетривиальным. А так как эвристические модели в основном нелинейные, то построения корректного алгоритма становится практически неразрешимой задачей. Поэтому возникает проблема понижения требований к эвристике  $\mathfrak{M}$ . Такая и некоторые другие проблемы решаются в моделях, основанных на принципе корректировки [Zhuravlev, 2010].



---

**Библиография**

---

- [Журавлев, 1978] Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. — 1978. — Т. 33. — С. 5–68.
- [Рудаков, 1989] Рудаков К.В. Об алгебраической теории универсальных и локальных ограничений для задач классификации // Распознавание, классификация, прогноз. М.: Наука, 1989
- [Krasnoproshin, 2006] V.V.Krasnoproshin V.A.Obraztsov Problem of Solvability and Choice of Algorithms for Decision Making by Precedence // Pattern Recognition and Image Analysis. 2006. Vol. 16. no 2.- p.p.155-169.
- [Zhuravlev, 2010] Yu. I. Zhuravlev, S. V. Ablameiko, A. S. Biryukov, A. A. Dokukin, V. V. Krasnoproshin, V. V. Obraztsov, M. Yu. Romanov, and V. V. Ryazanov Algorithms for Algebraic and Logical Correction and Their Applications // Pattern Recognition and Image Analysis. 2010. Vol. 20. no 2.- p.p.105-117.

---

**Информация об авторах**

---

**Виктор Краснопрошин** – заведующий кафедрой МО АСУ, ФПМИ, Белорусский государственный университет, пр-т Независимости, 4, Минск, 220050, Беларусь; e-mail: krasnoproshin@bsu.by

**Владимир Образцов** – доцент кафедры МО АСУ, ФПМИ, Белорусский государственный университет, пр-т Независимости, 4, Минск, 220050, Беларусь; e-mail: obraztsov@bsu.by

## МЕТОД ВЫДЕЛЕНИЯ ПРИЗНАКОВ НА ОСНОВЕ СОВМЕСТНЫХ ВЕКТОРОВ МОМЕНТНЫХ ФАЗОВЫХ РАСПРЕДЕЛЕНИЙ И МОМЕНТОВ ЗЕРНИКЕ

А.Н. Чернодуб

**Аннотация:** *Предлагается новый метод получения векторов признаков на основе инвариантных моментов называемых векторами моментных фазовых распределений для решения задачи распознавания двумерных полутоновых изображений. Приводятся данные численных экспериментов по классификации двумерных изображений, имеющих разный размер и меняющийся угол поворота относительно фронтальной оси. Исследуется устойчивость предлагаемого метода к шуму.*

**Ключевые слова:** *распознавание образов, обработка изображений, теория моментных инвариантов, полиномы Зернике*

**ACM Classification Keywords:** *1.4 Image Processing and Computer Vision, 1.4.7 Feature Measurement, Moments, Invariants.*

---

### I. Введение

Теория моментных инвариантов для двумерных изображений возникла с выходом работы М.К. Ху [Hu, 1962], в которой предложен способ вычисления векторов признаков для полутоновых изображений, не меняющихся при произведении аффинных преобразований над изображением: изменении размера, повороте и сдвиге. Была доказана теорема уникальности, гарантирующая взаимнооднозначное соотношение между исходными изображениями и генерируемыми векторами признаков, “геометрическими инвариантными моментами” (“geometric moments invariants”). Они с успехом применялись для распознавания печатного текста, классификации силуэтов кораблей и самолетов. Позже, М. Тиг [Teague, 1980] предложил способ вычисления инвариантных к аффинным преобразованиям моментов путем разложения функции интенсивности изображения по ортогональным полиномам Зернике. Было доказано, что моменты Зернике сводятся к геометрическим моментам, также была выведена формула, по которой моменты Зернике могут быть получены из геометрических моментов, и наоборот. Продолжением было опробование других множеств полиномов в качестве базисных функций. Ч. Х. Тех и Р. Т. Чин предложили в качестве базисных функций использовать ортогональные полиномы псевдо-Зернике и полиномы Лежандра [Teh & Chin, 1988], Ю. Шенг и Л. Шен ввели ортогональные моменты Фурье-Меллина [Sheng & Shen, 1994], Р. Мукундан и др. предложили моменты на основе полиномов Чебышева [Munkundan et. al., 2001], П.Т. Яп и др. предложили использовать моменты на основе полиномов Кравчука [Yap et. al, 2003], Х.К. Жу и др. ввели в обращение моменты Рака [Zhu et. al., 2007].

В нашей работе мы предлагаем новый способ вычисления векторов признаков на основе теории инвариантных моментов, называемый моментными фазовыми распределениями, который подходит для всех Зернике-подобных моментов. Будучи добавленными к вычисленным обычным способом моментам, моментные фазовые распределения способны улучшать качество распознавания объектов двумерных изображений при небольшом собственном размере, а также незначительных требованиях к дополнительным вычислительным ресурсам.

---

### II. Инвариантные моменты Зернике

Несмотря на свою долгую историю, моменты Зернике остаются очень популярными. На основе моментов Зернике создаются практические приложения по распознаванию образов, а их свойства продолжают активно исследоваться [Zhenjiang, 2000], [Rodtook & Makhanov, 2005], [Rouze et al., 2006], [Kim & Kim, 2008]. В связи с этим, мы используем именно моменты Зернике для описания предлагаемого способа вычисления дополнительного вектора признаков, моментных фазовых распределений.

Рассмотрим схему вычисления моментов Зернике. Пусть черно-белое полутоновое изображение на плоскости описывается функцией интенсивности  $f(x, y)$ , имеющей область значений на отрезке  $[0, 1]$ . Значение "0" кодирует абсолютно черный цвет, значение "1" – абсолютно белый. Промежуточные значения между 0 и 1 соответствуют различным оттенкам серого. На рис. 1, слева, показана функция интенсивности для белого символа "А", находящегося на черном фоне.

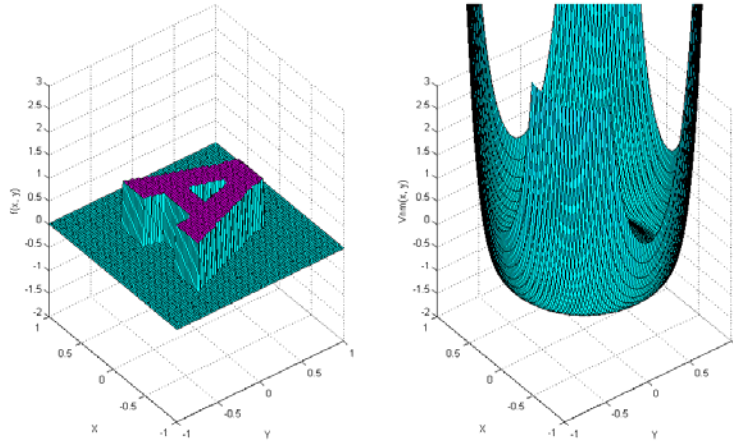


Рис. 1. Представление изображения символа "А" в виде функции интенсивности изображения  $f(x, y)$  (слева), пример базисной функции  $V_{6,4}$ , по которой проводится разложение полутонового изображения (справа).

Для обеспечения инвариантности к сдвигу, начало координат для вычисления моментов всегда устанавливается в точку "центра масс" изображения. Точка центр масс определяется по формуле:

$$(x_c; y_c) = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (1)$$

де  $m_{pq}$  — геометрические моменты по М.К. Ху [Ху, 1962]:

$$m_{pq} = \iint x^p y^q f(x, y) dx dy \quad (2)$$

Величину  $m_{00}$  можно интерпретировать как полную "массу" изображения.

Моменты Зернике  $ZM_{nm}$ , определенные на единичном круге  $0 \leq \sqrt{x^2 + y^2} \leq 1$ , вычисляются по формуле [Abdallah, 2000], для дискретного случая:

$$ZM_{nm} = \frac{n+1}{\pi} \sum_{x=0}^N \sum_{y=0}^M f(x, y) [V_{nm}(x, y)]^* \quad (3)$$

$$V_{nm}(x, y) = Z_{nm}(x, y) \exp(i \arctg(y/x)) \quad (4)$$

$$Z_{nm}(x, y) = \sum_{k=0}^{(n-m)/2} (-1)^k \frac{(n-k)!}{k! \left(\frac{n+m}{2} - k\right)! \left(\frac{n-m}{2} - k\right)!} (x^2 + y^2)^{(n-2k)/2} \quad (5)$$

где:  $n$  – порядок момента,  $m$  – порядок гармоники,  $n \in N^+$ ,  $m \in N$ ,  $|m| \leq n$ ,  $(n - |m|)$  четное, символом \* обозначается комплексное сопряжение. Здесь  $ZM_{nm}$  – комплексные моменты Зернике,  $Z_{nm}$  –

полиномы Зернике,  $V_{nm}$  - базисные функции, по которым происходит разложение изображения. На рис. 1, справа показан пример базисной функции  $V_{nm}$  при  $n = 6$ ,  $m = 4$ . Модули векторов моментов Зернике инвариантны к повороту:

$$|ZM^{ROTATED}_{nm}| = |ZM_{nm}| \quad (6)$$

Векторы признаков Зернике порядка  $N$  формируются согласно формуле:

$$ZMI_N = \left( \frac{|ZM_{11}|}{m_{00}}, \frac{|ZM_{20}|}{m_{00}}, \frac{|ZM_{22}|}{m_{00}}, \dots, \frac{|ZM_{NN}|}{m_{00}} \right)^T \quad (7)$$

Каждая компонента вектора признаков делится на массу изображения  $m_{00}$ , которая пропорциональна его размеру. Этим обеспечивается инвариантность объекта к изменению масштаба. Заметим, что  $m_{00} \equiv ZM_{00}$ . На практике, для распознавания образов редко используются векторы признаков порядка больше 15 в силу меньшей устойчивости моментов более высокого порядка к шуму, что снижает точность распознавания.

### III. Восстановление ориентации изображения по фазам моментов

При вращении объекта во фронтальной плоскости, комплексный вектор момента Зернике также поворачивается вместе с ним. При значении порядка гармоники момента  $m = 1$ , комплексный вектор выполняет один полный оборот относительно начала координат за время одного полного оборота объекта, при  $m = 2$  – вектор момента делает два полных оборота за время одного полного оборота объекта и т.д. Значения фаз моментов, то есть углов поворота векторов моментов на комплексной плоскости относительно вектора  $(1, 0)$  умышленно не используются, поскольку в качестве признаков берут длину комплексного вектора. Тем не менее, их можно получить из значений моментов.

Пусть  $m = 1$ . Тогда, фазу  $\alpha \in [0; 2\pi]$  для момента  $ZM_{n,1}$  порядка  $n$  можно вычислить по формуле:

$$\left\{ \begin{array}{l} \text{if } \operatorname{Re}(ZM_{n,1}) > 0, \quad \alpha = \operatorname{arctg} \frac{\operatorname{Im}(ZM_{n,1})}{\operatorname{Re}(ZM_{n,1})} + \frac{3\pi}{2} \\ \text{if } \operatorname{Re}(ZM_{n,1}) < 0, \quad \alpha = \operatorname{arctg} \frac{\operatorname{Im}(ZM_{n,1})}{\operatorname{Re}(ZM_{n,1})} + \frac{\pi}{2} \\ \text{if } \operatorname{Re}(ZM_{n,1}) = 0, \quad \alpha = 0 \end{array} \right. \quad (8)$$

В своей диссертации [Abdallah, 2000] С. Абдалла показал, что информацию о фазах моментов можно использовать для дополнительного повышения качества классификации объектов, поскольку по нему можно восстанавливать исходную ориентацию изображения и использовать ее как обратную гипотезу. Этот метод работает, если в наличии есть дополнительная априорная информация о возможных ограничениях в ориентации объекта. Например, для стационарной системы видео-слежения с блоком распознавания лиц можно ожидать невозможность наличия в поле зрения камеры перевернутых «вверх ногами» лиц. Этот метод не получил большого распространения из-за необходимости указания допустимых областей для фаз моментов различных образов вручную или использования примеров изображений при всех допустимых поворотах для автоматического выявления допустимых областей фаз моментов классификатором, что не всегда является возможным.

#### IV. Моментные фазовые распределения и формирование вектора признаков на их основе

Рассмотрим предлагаемый метод получения моментных фазовых распределений. Пусть на единичном круге задано полутоновое изображение. Определим количество круговых сегментов  $L$ , на которое мы разбиваем изображение,  $L \geq 3$ ,  $L \in \mathbb{Z}$ . Для каждого  $i$ , такого что  $1 \leq i \leq L$ , определим соответствующий круговой сегмент  $R_i$  по формуле:

$$R_i = \forall x, y : r_{i-1} < \sqrt{x^2 + y^2} \leq r_i, \quad r_i = \frac{1}{L} \times i \quad (9)$$

При  $i = 1$ , круговой сегмент является собой круг радиуса  $\frac{1}{L}$  с центром в точке начала координат, которая выбирается в точке центра масс изображения, а при  $2 \leq i \leq L$  круговой сегмент представляет собой кольцо ширины  $\frac{1}{L}$  с тем же центром. Определим фазы  $\alpha_i$  для каждого кругового сегмента  $R_i$  по формуле (9). Порядки моментов, которые используются для вычисления фаз, должны быть строго зафиксированы. На рис. 2 показано разделение изображения на три круговых сегмента по описанному выше способу и возможные фазы моментов для каждого из сегментов при параметре  $L = 3$ .



Рис. 2. Фазы моментов  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ , вычисленные для разных круговых сегментов  $R_1$ ,  $R_2$ ,  $R_3$  оригинального изображения и их поведение в случае поворота изображения вокруг фронтальной оси

По построению круговых сегментов, при вращении изображения в фронтальной оси, а также при изменении масштаба изображения и трансляции, ориентация фаз моментов относительно друг друга не меняется, углы между комплексными векторами моментов остаются постоянными. Относительное угловое расстояние  $d_k \in [0; \pi]$  между двумя последовательно расположенными моментами  $\alpha_k$  и  $\alpha_{k+1}$  может быть вычислено по формуле:

$$d_k = \sqrt{\cos(\alpha_{k+1} - \alpha_k)^2 + \sin(\alpha_{k+1} - \alpha_k)^2} \quad (10)$$

для  $1 \leq k \leq K$ ,  $K = L - 1$ . Моментные фазовые распределения  $PD_K$  порядка  $K$  вычисляются по формуле:

$$PD_K = (d_1; \dots; d_K)^T \quad (11)$$

Они могут быть использованы в качестве самостоятельных векторов признаков, а могут добавляться к векторам признаков на основе моментов Зернике, вместе формируя совместные векторы признаков:

$$ZPDMI_{N,K} = \left( \frac{|ZM_{11}|}{m_{00}}, \frac{|ZM_{20}|}{m_{00}}, \frac{|ZM_{22}|}{m_{00}}, \dots, \frac{|ZM_{NN}|}{m_{00}}; d_1; \dots; d_K \right)^T \quad (12)$$

Векторы признаков, вычисленные по последнему способу, представляют особый интерес. Во-первых, они содержат в себе моменты Зернике, эффективность которых в области распознавания двумерных образов была подтверждена многими практическими исследованиями. Во-вторых, в них содержатся новые компоненты, соответствующих моментным фазовым распределениям. Эти новые признаки несут в себе дополнительную информацию о структуре изображения, имеют небольшой объем и могут быть эффективно вычислены.

## V. Результаты экспериментов по классификации изображений

Мы проводили тестирование эффективности работы предлагаемого метода выделения признаков на задаче классификации изображений, взятых из базы полутоновых изображений COIL-20 [Nene et. al., 1996]. Выборка для экспериментов была подготовлена по способу, идентичному использованному в работе [Zhang et. al., 2009]: из базы данных было выбрано 8 изображений, показанных на рис. 3.

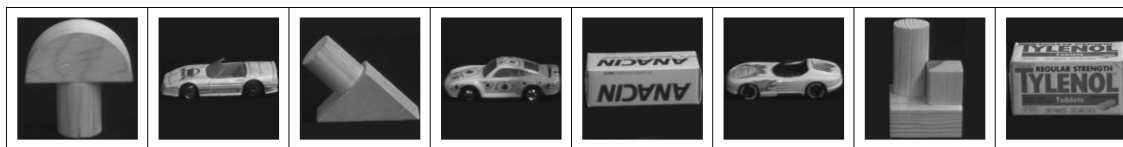


Рис. 3. Оригинальные изображения из базы данных Columbia Database, используемые для экспериментов по классификации изображений

Причиной выбора именно этих примеров была их большая внешняя схожесть. Оригинальный размер изображений – 128x128. Затем каждое из изображений было программно сжато/растянуто с коэффициентами масштабирования  $\lambda \in \{0.5, 0.75, 1, 1.5, 2\}$  и повернуто вокруг фронтальной оси с на углы  $\beta \in \{30^\circ, 60^\circ, \dots, 360^\circ\}$ , формируя таким образом множество из 480 изображений.

В качестве классификатора был использован простейший классификатор, метод k ближайших соседей при  $k = 1$ . Для обучения классификатора использовались только оригинальные изображения, показанные на рис. 3, по одному для каждого класса. Все остальные изображения были использованы для оценки качества классификации. Точность классификации  $\eta$  вычислялась по формуле:

$$\eta = \frac{\text{Количество правильно классифицированных изображений}}{\text{Общее количество изображений}} \times 100\% \quad (13)$$

В ходе первого эксперимента, в качестве векторов признаков использовались векторы, содержащие только моментные фазовые распределения при количестве круговых сегментов  $L = 3$ , определяемые формулой (11). Результаты эксперимента приведены в табл. 1.

Табл. 1. Точность классификации полутоновых изображений при использовании векторов признаков, содержащих только моментные фазовые распределения Зернике в зависимости от параметров.

	n=1, m=1	n=3, m=1	n=5, m=1	n=7, m=1	n=9, m=1	n=11, m=1	n=13, m=1	n=15, m=1
PD, точность классификации, %	45.975	67.373	59.958	61.017	58.686	52.331	46.61	33.686

Точность классификации получилась невысокой, что можно объяснить небольшим размером векторов признаков,  $K = 2$ . Использовать такие векторы признаков сами по себе в случае применения простых классификаторов нельзя.

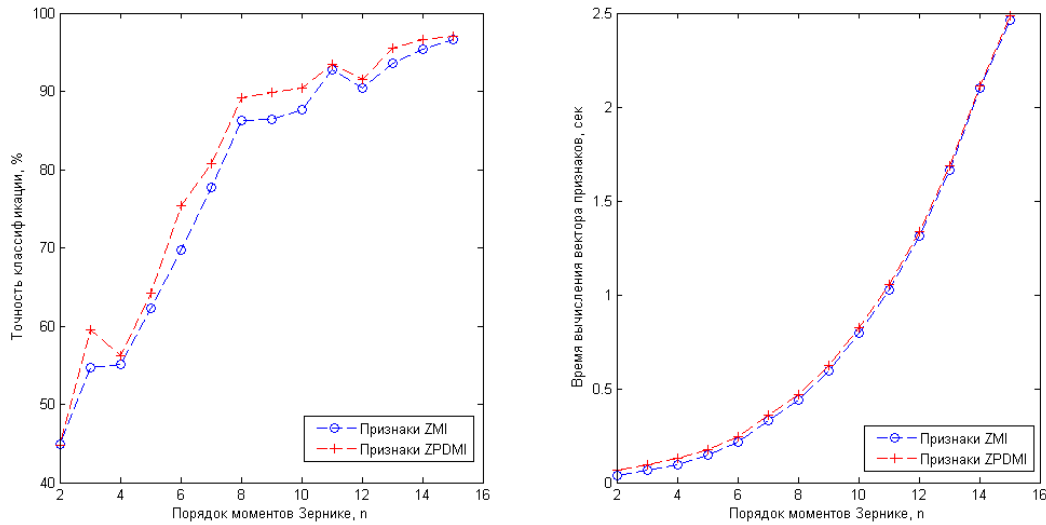


Рис. 4. Точность классификации изображений в зависимости от порядка используемых моментных функций (слева), среднее время вычисления вектора признаков в зависимости от порядка моментных функций (справа).

Во втором эксперименте мы сравнивали точность классификации, полученную при использовании традиционных инвариантных моментов Зернике ZMI, вычисленных по формуле (7) и предлагаемых совместных векторов моментных фазовых распределений и моментов Зернике ZPDMI, полученных по формуле (12), для различных порядков моментов Зернике, входящих в оба вида векторов. Для вычисления моментных фазовых распределений всегда использовались параметры  $n=1$ ,  $m=1$ .

Как видно из графиков на рис. 4, точность классификации, полученная при использовании предлагаемых векторов признаков ZPDMI получилась лучше, чем точность классификации для традиционных признаков ZMI для моментов всех порядков. Разница во времени вычисления векторов признаков при этом возросла незначительно.

В третьем эксперименте, наше множество из 480 изображений было искусственно зашумлено помехами типа "salt & pepper", с уровнем шум/сигнал 1%, 2%, 4% и 10%. Результаты классификации при использовании зашумленных изображений приведены в табл. 2.

Табл.2. Точность классификации изображений с различным уровнем шума для различных методов

	Без шума	Шум 1%	Шум 2%	Шум 4%	Шум 10%
Традиционный метод ZMI, % правильных ответов	96.61	96.19	94.07	95.76	85.17
Предлагаемый метод ZPDMI, % правильных ответов	97.03	96.82	94.28	95.97	86.44

Как видно из последней таблицы, предлагаемый метод формирования векторов признаков обеспечивает большую устойчивость к шуму, чем традиционный метод моментов Зернике.

## VI. Выводы

Предложенный нами метод выделения признаков, основан на идее вычисления взаимного распределения фаз моментов, полученных для различных круговых сегментов изображения, которые рассматриваются как независимые изображения. Эта идея не является новой. Распределение ориентаций градиентов для блоков из пикселей используется для выделения признаков в популярных методах препроцессинга изображений SIFT [Lowe, 2004], SURF [Bay et. al, 2006], CHoG [Chandrasekhar et. al, 2009] и др. Разделение единого изображения на несколько независимых сегментов и выделение признаков каждого из них предлагался для инвариантных моментов псевдо-Зернике [Kanana et. al., 2008].

В нашей работе мы вычисляем фазы моментов для специальным образом определенных круговых сегментов изображения, а затем используем их взаимные ориентации, которые являются инвариантными к повороту, масштабированию и трансляции. Было экспериментально показано, что новые объединенные векторы признаков моментных Зернике и моментных фазовых распределений обеспечивают лучшее качество классификации, чем стандартные моменты Зернике, а их вычисление не отнимает дополнительно большого количества ресурсов.

Требуют дополнительного исследования вопросы установления оптимального количества используемых круговых сегментов, оптимального порядка моментов для вычисления значений фаз моментов, влияния пересечения круговых сегментов на качество распознавания. Кроме того, открытым остается вопрос о возможности использования чистых векторов моментных фазовых распределений в качестве генераторов слабых признаков совместно с многомодульными классификаторами, работающими по схемам бустинга. Такие признаки могли бы стать потенциальными заменителями неинвариантных к аффинным преобразованиям вейвлетов Хаара в детекторах объектов, построенных по типу детекторов лиц Виолы и Джонса [Viola & Jones, 2001]. Заслуживает также дополнительной проверки эффективность предлагаемого метода применительно к инвариантным моментам других типов.

## Список литературы

- [Abdallah, 2000] S. M. Abdallah. Object Recognition via Invariance. In: Ph.D. thesis, The University of Sydney, 2000.
- [Hu, 1962] M.K. Hu. Visual pattern recognition by moment invariants. In: IRE Transactions on Information Theory 8, 1962, pp. 179–187.
- [Bay et. al, 2006] H. Bay, T. Tuytelaars, L. V. Gool. SURF: Speeded Up Robust Features. In: Lecture Notes in Computer Science, Computer Vision – ECCV 2006, Volume 3951/2006, pp. 404-417.
- [Chandrasekhar et. al, 2009] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, B. Girod. CHoG: Compressed Histogram of Gradients: A Low Bit-Rate Feature Descriptor. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'2009, pp.2504-2511.
- [Flusser et.al., 2009] J. Flusser, T. Suk, B. Zitová. Moments and Moment Invariants in Pattern Recognition. Wiley & Sons Ltd., 2009.
- [Kanana et. al., 2008] H. R. Kanana, K. Faeza, Y. Gaob. Face recognition using adaptively weighted patch PZM array from a single exemplar image per person. In: Pattern Recognition 41, 2008, pp. 3799 – 3812.
- [Lowe, 2004] Lowe, D. G.. Distinctive Image Features from Scale-Invariant Keypoints. In: International Journal of Computer Vision, 60, 2, 2004, pp. 91-110.
- [Munkundan et. al., 2001] R. Mukundan, S.H. Ong, P.A. Lee. Image analysis by Tchebichef moments. IEEE Trans Image Processing 10, 2001, pp. 1357–1364.
- [Kim & Kim, 2008] H.J. Kim and W. Yul. Kim. Eye Detection in Facial Images Using Zernike Moments with SVM, ETRI Journal, Vol. 30, Num. 2, April 2008, pp. 335-337.
- [Nene et. al., 1996] S. A. Nene, S. K. Nayar and H. Murase. Columbia Object Image Library (COIL-20). In: Technical Report CUCS-005-96, February 1996. Available in World Wide Web at: <http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php>
- [Rodtook & Makhanov, 2005] S. Rodtook, S.S. Makhanov. Numerical experiments on the accuracy of rotation moments invariants. In: Image and Vision Computing 23 (2005) 577–586.



- [Rouze et al., 2006] N. C. Rouze, V.C. Soon, G.D. Hutchins. On the connection between the Zernike moments and Radon transform of an image. In: Pattern Recognition Letters 27 (2006), pp. 636–642.
- [Shu, 2007] H. Shu, L. Luo and J.-L. Coatrieux. Moment-based approaches in imaging. Part 1, basic features. IEEE Eng Med Biol Mag. 2007 Sep–Oct; 26(5): 70–74.
- [Sheng & Shen, 1994] Y. Sheng and L. Shen. Orthogonal Fourier-Mellin moments for invariant pattern recognition. In: J. Opt. Soc. Am. A, Vol. 11, No. 6, June 1994.
- [Teague, 1980]. M. Teague. Image analysis via the general theory of moments. In: Journal Optical Society America 70 (1980) 920–930.
- [Teh & Chin, 1988]. C.H. Teh, R.T. Chin. On image analysis by the method of moments. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (1988) 496–513.
- [Viola & Jones, 2001] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. In: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 2001.
- [Wallin & Kibbler, 1995] A. Wallin, O. Kibbler. Complete Sets Of Complex Zernike Moment Invariants and the Role of the Pseudoinvariants. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 11, Nov. 1995.
- [Yap et. al, 2003] P.T. Yap, R. Paramesran, S.H. Ong. Image Analysis by Krawtchouk Moments. IEEE Trans Image Processing 12, 2003, pp. 1367–1377.
- [Zhu et. al, 2007] H.Q. Zhu, H.Z. Shu, J. Liang, L.M. Luo, J.L. Coatrieux. Image analysis by discrete orthogonal Racah moments. Signal Processing 87, 2007, pp. 687–708.
- [Zhenjiang, 2000] M. Zhenjiang. Zernike moment-based image shape analysis and its application. In: Pattern Recognition Letters 21 (2000) pp. 169-177.
- [Zhang et. al., 2009] H. Zhang, H.Z. Shu, P. Haigron, B.S. Li, L.M. Luo. Construction of a complete set of orthogonal Fourier-Mellin moment invariants for pattern recognition applications. In: Image and Vision Computing, Volume 28, Issue 1, 2010, pp. 38-44.

---

### Информация об авторах

---



**А.Н. Чернодуб** – аспирант Института Программных Систем НАНУ, работает в Институте Проблем Математических Машин и Систем НАНУ. Адрес: г. Киев, ул. академика Глушкова 42, ИПММС НАНУ, отдел Нейротехнологий; e-mail: [achernodub@immsp.kiev.ua](mailto:achernodub@immsp.kiev.ua)

Научные интересы: интеллектуальная обработка изображений, биометрическая идентификация, нейронные сети, нейроуправление.

## ОБУЧЕНИЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ МЕТОДОМ КОНТРОЛИРУЕМОГО ВОЗМУЩЕНИЯ ДЛЯ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ ОБЪЕКТАМИ

Д. А. Дзюба, А. Н. Чернодуб

**Аннотация:** Предложен новый метод нейруправления с применением нелинейного нейроконтроллера на основе рекуррентного прерцептрона, обучаемого в режиме реального времени с использованием метода контролируемого возмущения управления динамическим объектом. Обучение нейроконтроллера осуществляется без использования нейромулятора, предварительно обученного прямой или инверсной динамике объекта управления. Приводятся данные численных экспериментов по сравнению качества нового метода нейруправления и традиционного ПИД-управления на примере решения задачи стабилизации перевернутого маятника.

**Ключевые слова:** нейруправление, рекуррентные нейронные сети, нелинейные динамические объекты, инверсная динамика, стабилизация перевернутого маятника, обучение с подкреплением.

**ACM Classification Keywords:** I.2.8 Problem Solving, Control Methods, and Search - Control theory, I.2.9 Robotics - Autonomous vehicles, I.2.6 Learning - Connectionism and neural nets.

### 1. Вступление

Применение искусственных нейронных сетей для создания обучаемых систем управления динамическими объектами является перспективным направлением в теории управления, теории искусственного интеллекта, нейрофизиологии [Омату и др., 2000]. Нейронные сети обладают многими интригующими свойствами, которые делают их мощным инструментом для создания управляющих систем: способность к обучению на примерах, способность к гладкой интерполяции и экстраполяции данных на ранее не виденных нейросетью примерах, возможность синтеза нелинейных контроллеров, способность адаптации к изменяющимся свойствам объекта управления и внешней среды в режиме реального времени, большая по сравнению с классической фон-Неймановской архитектурой устойчивость к повреждениям своих элементов в силу изначально заложенного в нейросетевую архитектуру параллелизма. Среди ранних примеров систем нейруправления можно назвать работу Б. Видроу [Widrow, 1986]. Он применил линейные нейронные сети с линией задержек на входе для синтеза системы управления, состоящей из двух модулей, один из которых был обучен инверсной динамике объекта и работал как прямой контроллер без обратной связи, а второй выполнял функцию восстановления исходного сигнала из зашумленного сигнала на входе объекта. Минусами этой схемы является ее неспособность работать с нестабильными объектами, необходимость наличия точной математической или имитационной модели объекта для обучения модуля фильтрации шума. В работе Д. Псалтиса и А. Сидериса [Psaltis & Sideris, 1988] в качестве нейросети использован статический многослойный перцептрон с линией задержек, что позволяет синтезировать нелинейный контроллер, позволяющий обеспечивать повышенное качество управления при работе с нелинейными объектами управления, все остальные минусы схемы Б. Видроу остались неразрешенными. В этой же работе Д. Псалтис и др. предложили метод специализированного инверсного управления, в котором контроллер на основе многослойного перцептрона с линией задержек обучался управлять объектом на основе ошибки между реальным выходом объекта и целевым выходом (уставкой). Для применения этой схемы необходимо знание якобиана объекта управления, либо аппроксимация формирующих якобиан частных производных объекта управления по значениям его входов и выходов, на основе использования аппарата вычислительной математики. В другой схеме нейруправления, предложенной независимо К.С. Нарендрой и К. Пасарати [Narendra & Pasarchy, 1990], Б. Вербосом [Werbos, 1990] и Джорданом и Румельхартом [Jordan & Rumelhart, 1990], и получившей впоследствии значительную популярность, для получения ошибки нейроконтроллера вместо

непосредственного вычисления якобиана объекта управления используется механизм обратного распространения ошибки через прямую нейросетевую модель объекта управления, предварительно обученного точно воспроизводить динамику поведения объекта управления. Также в работе [Narendra & Pasarchy, 1990], было указано, что для эмуляции поведения динамического объекта наилучшей моделью нейронных сетей являются рекуррентные многослойные перцептроны, позже Х. Сигельман и др. доказали это строго [Siegelmann, 1997].

В описанных методах нейроуправления с обучением нейроконтроллера по ошибке отклонения объекта управления от целевой траектории используется мгновенное значение ошибки на каждом такте работы системы. Нейроконтроллер на этапе функционирования выполняет действия, ориентированные на минимизацию именно текущей ошибки, что может дать худшее интегральное качество управления в долгосрочной перспективе. Для преодоления этого недостатка, были предложены методы нейросетевого прогнозирующего управления [Hagan & Demuth, 1999] и системы адаптивных критиков [Prokhorov & Wunsch, 1997]. В схеме нейросетевого прогнозирующего управления также производится обучение нейроэмулятора прямой динамике объекта, который затем используется для предсказания поведения объекта на несколько шагов вперед. При этом пробуются разные управляющие воздействия, и выбирается то, которое дает наименьшую интегральную ошибку. Перебор различных управляющих воздействий осуществляет специальный оптимизационный модуль в составе системы управления. Работа систем адаптивных критиков основана на применении принципа Беллмана для выбора контроллером управляющего действия, оптимального в смысле определения наилучшей стратегии. В них выделяется отдельный нейросетевой модуль на основе многослойного перцептрона, называемый «критик», который обучается по наблюдению за реальным поведением системы возвращать оценку прогнозируемой ошибки управления на некоторое количество шагов вперед, в зависимости от поданного на вход управляющего воздействия. Обучение контроллера, входящего в систему управления адаптивного критика, выполняется по принципу минимизации отклика возвращаемого критиком. В ходе обучения системы, оба модуля обучаются попеременно: критик обучается выполнять более качественные оценки (итерация по значениям), а контроллер обучается осуществлять более правильные действия, ведущие к лучшим оценкам (итерация по стратегиям). В случае, когда критик прогнозирует ошибку только на один шаг вперед, схема управления на основе адаптивной критики фактически вырождается в схему с обратным распространением ошибки через прямой нейроэмулятор. Кроме этих схем, также были предложены схемы обучения нейронной сети для корректировки коэффициентов ПИД-контроллера [Ruano et. al., 1992], схемы по использованию нейроконтроллеров параллельно с существующими классическими ПИД-контроллерами [Kawato et. al., 1988], нейросистемы на основе многомодульных самоорганизующихся нейросетей [Wolpert & Kawato, 1998], [Ronco et. al, 1996] и др.

Наблюдаемая тенденция к усложнению нейроуправления отражает стремление возможно более точно представить инверсную модель нелинейного объекта управления, используемую при обучении нейроконтроллера. Модель на основе прямого нейроэмулятора с механизмом обратного распространения ошибки дает возможность учитывать нелинейность объекта и фокусировать управление в локальной области, допускающей линейную аппроксимацию его инверсной динамики. Однако создание эффективного нейроэмулятора для объекта со сложным нестационарным поведением часто проблематично. В таких случаях информацию, необходимую для обучения нейроконтроллера получают с использованием более сложных методов типа прогнозирующего нейроуправления, адаптивной критики и т.п. Однако, всегда ли оправдано такое усложнение нейроуправления?

Мы предлагаем новую схему нейроуправления, в которой нейроконтроллер, основанный на рекуррентной нейронной сети, обучается управлению оперируя непосредственно объектом в режиме реального времени. Мы используем контролируемое малое возмущение управления, реакция на которое служит для определения направления коррекции работы контроллера. В нашей схеме отсутствует нейроэмулятор, а направление коррекции управления оценивается аналитически на каждом шаге обучения системы, что в итоге приводит к формированию инверсной модели объекта управления непосредственно в

нейроконтроллере. Для оценки эффективности предлагаемой схемы проводится экспериментальное сравнение ее работы с ПИД-управлением на примере задачи стабилизации положения перевернутого маятника.

## 2. Метод контролируемого возмущения управления

Для решения задачи управления с помощью нейроконтроллера необходимо предоставить сети либо пример оптимального управления (которое в ряде случаев неизвестно), либо знак производной от качества управления по управляющему сигналу для коррекции текущего выхода сети (обучение с подкреплением). Именно решению второй задачи и посвящена данная работа.

Для обучения с подкреплением используется информация о том, в какую сторону следует менять управление, чтобы его улучшить. По сути, это означает, что строится некоторая инверсная модель объекта управления, которая давала бы правильный знак производной от состояния объекта по управлению, и который можно было бы трансформировать в обучающий сигнал нейросети. Часто для получения такой инверсной модели используют нейроэмулятор, однако, поскольку нас интересует только знак, а не точная величина производной, можно воспользоваться более простым методом, а именно — определить его анализируя непосредственно поведение объекта. Если мы сравниваем состояние объекта после применения двух близких управлений — то, при соблюдении некоторых условий, мы можем в некотором приближении определить производную от состояния объекта по примененному управлению.

Найти производную от состояния по управлению можно следующим образом: пусть  $X_0$  — обобщенные координаты системы в некоторый момент времени,  $T$  — целевое положение. Применим на следующем временном шаге некоторое управление  $U_1 - h$ , где  $h$  — вектор с одной ненулевой компонентой, по которой берется производная, значение которой достаточно мало (критерий достаточности будет определен ниже), и обозначим полученные координаты как  $X_1$ , а на шаге после этого —  $U_1 + h$ , и обозначим полученные координаты как  $X_2$ . Теперь наша задача состоит в том, чтобы определить, какое из этих управлений в большей степени соответствует достижению целевого положения. Точный ответ мог бы дать эксперимент, в котором объект вернули из состояния 1 в состояние 0, применили второе управление, и сравнили полученные состояния с целевым. Однако, поскольку мы имеем дело с реальным объектом, а не его моделью, это невозможно. Тем не менее, эту задачу можно решить в некотором приближении. Нами было использовано первое приближение, в нем точка, в которой оказалась бы система, если бы в состоянии 0 к ней было приложено управление  $U_1 + h$ , определяется как:

$$X_1' = X_0 + (X_2 - X_1) + (X_0 - X_1) dt \quad (1)$$

здесь первый член — собственно исходное положение системы  $X_0$ , второй член — это фактическое смещение системы, полученное в результате применения второго управления, а третий член вводит поправку на то, что за счет применения первого управления, скорость системы изменилась.

Тогда производная от положения по некоторой компоненте управления может быть записана как

$$\frac{\partial X}{\partial U_i} = \frac{X_1' - X_1}{2|h|} \quad (2)$$

где индекс  $i$  соответствует номеру единственной ненулевой компоненты вектора  $h$ .

Теперь, когда у нас есть возможность сравнивать результаты различных управлений, мы можем применить следующий метод:

- На вход нейроконтроллера подается текущее состояние системы (координаты, скорости), и сеть генерирует некоторое управление  $U$ ;

- К системе последовательно применяется управление  $U+h$  и  $U-h$ , где номер ненулевой компоненты  $h$  последовательно пробегает все значения, с запоминанием полученных состояний;
- Для каждой пары полученных состояний системы вычисляется производная по соответствующей компоненте управления описанным выше способом;
- Выполняется один шаг обучения рекуррентной сети, где входом служит исходное состояние системы, а целевым значением — вектор  $U$ , смещенный согласно знаку производных в сторону приближения к целевому положению  $T$ .

Величина  $h$  выбирается достаточно малой, чтобы нелинейность системы слабо проявлялась на разнице между управлениями  $U+h$  и  $U-h$ , однако не слишком малой, поскольку она связана со скоростью обучения нейронной сети.

Описанный процесс — это один шаг обучения сети, повторив его достаточное число раз, можно научить рекуррентную нейронную сеть поддерживать заданное положение системы.

### 3. Модель перевернутого маятника

В качестве полигона для испытания системы нейруправления мы выбрали программную модель объекта «тележка с перевернутым маятником» в силу нескольких причин. Во-первых, этот объект является классическим объектом в теории управления [Michie & Chambers, 1968] и сейчас насчитывается более 2000 научных статей с описанием алгоритмов управления применительно к этому динамическому объекту [Стюарт и Норвиг, 2006], в том числе на основе как рекуррентных одномодульных [Wei et. al, 2001], так и нерекуррентных многомодульных [Ronco et. al, 1996] нейронных сетей. Во-вторых, это нелинейный динамический объект с сильно выраженной нелинейностью при отклонении маятника больше  $12^\circ$  от вертикальной оси [Ronco et. al, 1996], также он является сложным объектом для расчета управления на основе аппарата классической теории оптимального управления [Slotine & Li, 1991].

Схема перевернутого маятника показана на рис. 1. Тележка массы  $M$  может горизонтально перемещаться в одной степени свободы, вправо и влево. На тележке закреплен маятник массы  $m$  длины  $l$ , образующий угол  $\theta$  с вертикальной осью. В случае, когда  $\theta = 0$ , система находится в состоянии равновесия. По условию задачи, на систему действуют внешние силы, которые дестабилизируют ее, вынуждая маятник отклоняться от состояния равновесия. Если маятник отклоняется слишком далеко от положения равновесия, он падает вниз. Для преодоления этого, тележка оснащена мотором, на который может подаваться сила  $F$ , которая двигает тележку вместе с маятником вправо или влево.

Динамика перевернутого маятника описывается системой уравнений:

$$(M + m)x - ml\ddot{\theta} \cos \theta + ml\dot{\theta}^2 \sin \theta = F \quad (3)$$

$$ml(-g \sin \theta - x \cos \theta + l\ddot{\theta}) = 0 \quad (4)$$

Целью задачи контроллера является поддержание маятника в вертикальном положении в условиях внешних помех как можно более долгое время, получая на вход значение текущего угла отклонения  $\theta$  и генерируя значение управляющего воздействия  $F$ .

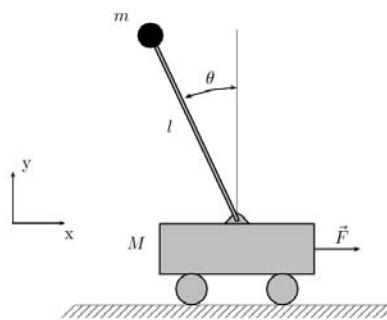


Рис. 1. Схема перевернутого маятника на тележке

Более качественной считается та система управления, которая может удерживать маятник на весу более долгое время. В ходе испытаний, для упрощения восприятия результатов экспериментов, проводился контроль только отклонения маятника, положение тележки в пространстве не фиксировалось. Так же было сделано, например, в работе [Ronco et. al, 1996].

#### 4. Экспериментальное исследование качества управления полученной нейросети

В нашей работе в качестве оценки качества управления мы использовали время в секундах, которое проходит от установки маятника в положение равновесия до выхода угла маятника за определенную границу при некотором уровне шума (в этой работе приведены результаты для порогового угла, равного примерно 10 градусам). Этот критерий позволяет эффективно сравнивать управление в ситуациях, когда маятник достаточно далек от положения равновесия, и нелинейные эффекты вносят ощутимый вклад, и достаточно популярен [Michie & Chambers, 1968], [Стюарт & Норвиг, 2006].

В целом эксперимент был построен следующим образом: модель маятника, импортированная из среды Simulink, приводилась в положение равновесия, потом к ней последовательно применялся один шаг случайного управления с заданной максимальной амплитудой (шум), и один шаг управления с помощью контроллера — рекуррентной сети, или PID, в зависимости от режима эксперимента. Если угол отклонения маятника не превысил заданный порог — шум и управление применялись повторно, и так до тех пор, пока угол отклонения маятника не достигал порогового значения. Время, которое прошло от момента установки маятника в положение равновесия до момента, когда отклонение превышало порог, записывалось как результат управления, маятник устанавливался в положение равновесия, и процесс повторялся снова. Схема управления показана на Рис. 2.

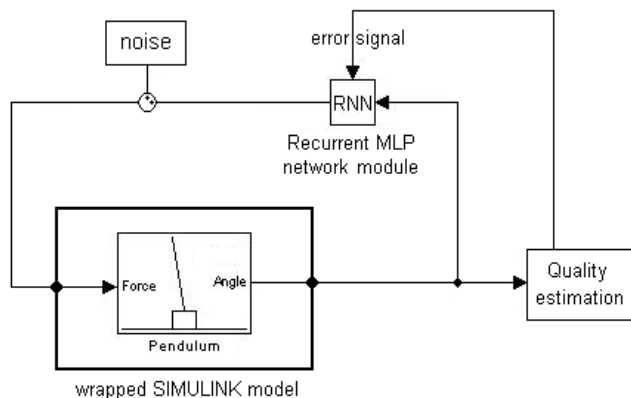
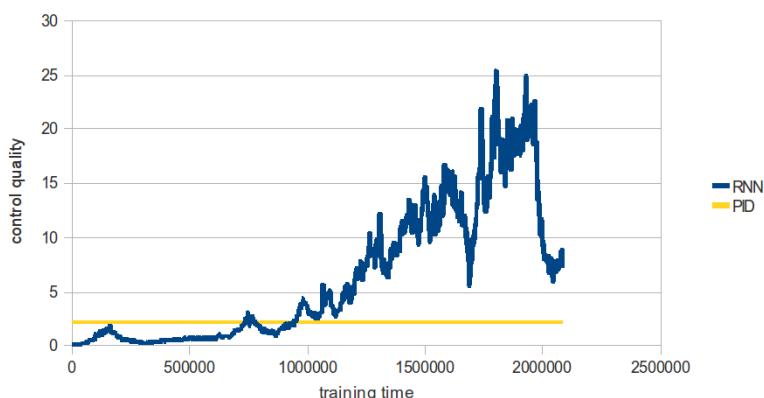


Рис. 2. Схема управления с помощью нейронной сети

Структура нейронной сети была следующей: на вход подавалось значение текущего угла, угловой скорости, и уставка по углу, эти данные попадали в линию задержек на 2 такта, выход сети попадал на вход через линию задержек в три такта. Таким образом сеть получала 9 внешних входов и 3 задержанных выхода, итого на вход попадало 12 чисел. Было проведено несколько экспериментов с разным числом нейронов в скрытом слое (от 3 до 12 нейронов). Эксперименты показали, что изменения числа скрытых нейронов в таких пределах дают не очень большие

изменения качества управления, и, поскольку задача оценки влияния этого параметра на эффективность управления не являлась предметом данного исследования, размер скрытого слоя был выбран равным 6 нейронам.



*Рис. 3. Зависимость качества управления от времени обучения рекуррентной сети. Для сравнения приведено качество управления для PID-контроллера*

Полученные данные изображены в виде графика на Рис. 3. Эксперименты показали, что качество управления рекуррентной сетью меняется немонотонно в процессе обучения, поэтому для оценки сеть использовалась в таком режиме:

Процесс обучения, описанный в разделе 2, повторялся 3000 раз, с установкой маятника в положение равновесия в начале серии повторов;

Сеть переключалась в режим управления, маятник устанавливался в положение равновесия, замерялось время до превышения прогового угла отклонения, замер повторялся 200 раз. Усредненная за это время характеристика качества управления использовалась как оценка сети на данном этапе обучения.

## 5. Выводы

Предложенный метод обучения рекуррентных сетей задачам управления может быть полезен в случаях, когда нет примеров оптимального управления некоторым объектом, но есть возможность оценить его состояние с точки зрения достижения некоторого целевого положения. Отсутствие в схеме нейромультипликатора позволяет существенно сократить время обучения, однако накладывает некоторые ограничения на объект управления — а именно, предложенная функция оценки знака производной состояния по управлению должна быть корректной для него.

Использованные нами рекуррентные сети имеют большой потенциал в подобных задачах, т.к. с одной стороны, рекуррентные связи обеспечивают отслеживание динамического состояния объекта, а с другой — нелинейность, присущая нейросетям, позволяет более эффективно управлять нелинейными объектами в сравнении с аналитически простыми линейными методами.

Обучаемые с помощью предлагаемого метода нейроконтроллеры могут адаптироваться к изменяющимся свойствам объекта управления, что делает их более выигрышными по сравнению с классическими неадаптивными контроллерами в тех случаях, когда параметры объекта меняются в ходе его функционирования.

Эти общие соображения полностью подтверждаются результатами экспериментов. Однако вопрос об оптимальной структуре сети и режиме обучения пока остается открытым.

## Список литературы

[Омату и др., 2000] С. Омату, М. Халид, Р. Юсоф. Нейроуправление и его приложения, пер. с англ. — М: ИПРЖР, 2000.

[Резник, 2009] А.М.Різник. Динамічні рекурентні нейронні мережі: Математичні Машины і Системи, 2009, №2, с.3-26.

- [Резник и Дзюба, 2010] А.М. Резник, Д.О.Дзюба. Динамічна автоасоціативна пам'ять, заснована на відкритій рекурентній нейронній мережі: Математичні Машини і Системи, 2010, №2, р. 45-51.
- [Стюарт & Норвіг, 2006] Р. Стюарт, П. Норвіг. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. — М. : Издательский дом "Вильямс", 2006.
- [Jordan & Rumelhart, 1990] M.I. Jordan and D.E. Rumelhart. Forward models: Supervised learning with a distal teacher. In: Cognitive Science, Vol. 16, pp.313- 355, 1990.
- [Hagan & Demuth, 1999] M.T. Hagan, H.B. Demuth. Neural Networks for Control. In: Proceedings of the 1999 American Control Conference, San Diego, CA, 1999, pp. 1642-1656.
- [Narendra & Pasarthi, 1990] K.S. Narendra, K. Parthasarathy K, Identification and control of dynamical systems using neural networks. In: IEEE Transactions on Neural Networks, 1, 1990, p. 4-27.
- [Kawato et. al., 1988] M.Kawato, Y. Uno, M. Isobe and R.Suzuki. Hierarchical neural network model for voluntary movement with application to robotics. In: IEEE Control Systems Magazine. Vol. 8, pp.8- 16, 1988.
- [Reznik & Dziuba, 2009] А.М. Reznik, D.A. Dziuba. Dynamic Associative Memory Based on Open Recurrent Neural Network. In: Proceeding of IJCNN'09, Atlanta, Georgia, USA, June 14-19, 2009.
- [Prokhorov & Wunsch, 1997] D. Prokhorov and D. Wunsch. Adaptive critic designs. In: IEEE Transactions on Neural Networks, 8(5), 1997, p. 997-1007.
- [Psaltis & Sideris, 1988] Demetri Psaltis, Athanasios Sideris, and Alan A. Yamamura. A Multilayered Neural Network Controller. In: IEEE Control Systems Magazine (1988), v. 8, Issue 2, p.17-21.
- [Ronco et. al, 1996] E. Ronco, J. Gawthrop, Y. Matter. Incremental Modular controllers network. In: Proceeding of the International Conference on Intelligent and Cognitive Systems (ICISC'96).
- [Ruano et. al., 1992] A. E. B. Ruano, P. J. Fleming, and D. I. Jones. Connectionist approach to PID tuning. In: IEEE Proceedings, Part D, 129:279-285, 1992.
- [Siegelmann, 1997] H.T. Siegelmann, B.G. Horne, and C.L. Giles. Computational capabilities of recurrent NARX neural networks. In: IEEE Trans. Systems, MAN, and Cybernetics - Part. B: Cybernetics, 1997, 27(2): 208-215.
- [Slotine & Li, 1991] J.-J. E. Slotine, W. Li. Nonlinear Control Systems Design. In: Prentice-Hall, 1991, p. 193.
- [Sutton & Barto, 1998] R.S. Sutton, A.G. Barto. Reinforcement Learning: An introduction. A Bradford book, 1998.
- [Michie & Chambers, 1968] D. Michie and Chambers. BOXES: An experiment in adaptive control. In: Dale E. and Michie D. (Eds.), Machine Intelligence 2, 1968, p. 125-133, Elsevier/North-Holland, Amsterdam, London, New York.
- [Wei et. al, 2001] W. Wei, W. von Seelen. Recurrent neuro-controller for an inverted pendulum using evolution strategy. In: International Journal of Systems Science, 2001, volume 32, number 5, pages 643-650.
- [Werbos, 1990] P. Werbos, Backpropagation through time: what it does and how to do it. In: Proc. IEEE, Vol. 78, No. 10, October 1990.
- [Widrow, 1986] Bernard Widrow. Adaptive Inverse Control. In: IFAC Adaptive Systems in Control and Signal Processing, Lund, Sweden, 1986.
- [Wolpert & Kawato, 1998] D.M. Wolpert, M. Kawato. Multiple Paired Forward and Inverse Models for Motor Control. In: Neural Networks, 1998, Vol. 11, Issue 7-8, pp. 1317 – 1329.

## Информация об авторах



**Д.А. Дзюба** — аспирант Института Программных Систем НАНУ, работает в Институте Проблем Математических Машин и Систем НАНУ. Адрес: г. Киев, ул. академика Глушкова 42, ИПММС НАНУ, отдел Нейротехнологий; e-mail: [ddziuba@immisp.kiev.ua](mailto:ddziuba@immisp.kiev.ua)  
 Научные интересы: динамическая ассоциативная память, нейроуправление, рекуррентные нейронные сети, интеллектуальная обработка изображений.



**А.Н. Чернодуб** – аспирант Института Программных Систем НАНУ, работает в Институте Проблем Математических Машин и Систем НАНУ. Адрес: г. Киев, ул. академика Глушкова 42, ИПММС НАНУ, отдел Нейротехнологий; e-mail: [achernodub@immisp.kiev.ua](mailto:achernodub@immisp.kiev.ua)  
 Научные интересы: интеллектуальная обработка изображений, биометрическая идентификация, нейронные сети, нейроуправление.



## ИССЛЕДОВАНИЕ МОДИФИКАЦИИ FOTSK НЕЧЕТКОЙ НЕЙРОННОЙ СЕТИ TSK В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ

Юлия Проскурня, Богдан Гривко

**Аннотация:** Существуют ситуации, когда нужно получить не только точечный прогноз, но еще и доверительный интервал к нему. Примером такой ситуации может выступить задача составления нечеткого инвестиционного портфеля, входными параметрами которой, в том числе, являются нижняя граница доходности, наиболее ожидаемое значение доходности и верхняя граница доходности ценных бумаг. Кроме того, доверительный интервал позволяет оценить точность прогноза.

В данной статье предложена модификация нечеткой нейронной сети TSK (Takagi, Sugeno и Kang'a), получившая название FOTSK (Fuzzy Output TSK), позволяющая получать на выходе нечеткое число вида: {нижняя граница; наиболее ожидаемое значение; верхняя граница}. Изменения затронули гибридный алгоритм обучения нечеткой нейронной сети TSK, а именно, его первый этап – расчет линейных параметров. В нечеткой нейронной сети FOTSK линейные параметры находятся из решения задачи линейного программирования, что, вдобавок, лишает нейронную сеть проблемы плохой обусловленности матрицы.

Также, авторами предложен алгоритм поиска оптимальной структуры нечеткой нейронной сети TSK (оптимального количества используемых нейронной сетью правил), который позволяет избежать перерегулирования.

В экспериментальных исследованиях было рассмотрено применение предложенной модификации FOTSK в задачах прогнозирования и произведено ее сравнение с оригинальной нечеткой нейронной сетью TSK, а также с НМГУА (Нечеткий Метод Индуктивного Моделирования). Исследования проводились в задачах прогнозирования курсов акций компаний IBM, Google, Apple, Coca-Cola и McDonald's на 28.05.2010. Выборки данных были сформированы по курсам акций вышеупомянутых компаний, взятых за период с 27.05.2009 по 27.05.2010.

**Ключевые слова:** нечеткие нейронные сети, прогнозирование, FOTSK, TSK.

**ACM Classification Keywords:** I.2.6 Learning – Connectionism and neural nets.

**Раздел конференции:** Neural and Growing Networks.

---

### Введение

Среди множества существующих методов прогнозирования особое место занимают нечеткие нейронные сети (ННС), а также методы индуктивного моделирования (МИМ), которые отличаются высоким качеством прогнозов и позволяют решать задачи принятия решений в условиях неопределенности.

Ярким представителем ННС и МИМ, соответственно, являются ННС TSK и нечеткий метод группового учета аргументов (НМГУА).

Однако выходом ННС TSK, в отличие от НМГУА, является четкое число (точечный прогноз), а в ряде случаев (напр., формирование нечеткого портфеля инвестора) необходимо иметь доверительный интервал, который характеризует точность прогноза. Для решения данной проблемы была разработана уникальная модификация ННС TSK с выходом модели в виде нечеткого числа, которая получила название FOTSK (Fuzzy Output TSK).

Целью данной статьи является рассмотрение и анализ модификации FOTSK ННС TSK в задачах прогнозирования, а также, сравнение результатов прогнозирования посредством данной модификации с результатами прогнозирования ННС TSK и НМГУА.

### Основные принципы и идея нечеткой нейронной сети TSK

Согласно [Зайченко, 2008] ННС TSK (Takagi, Sugeno и Kang'a) является обобщением нейронной сети ANFIS. Обобщенную схему вывода в модели TSK при использовании  $M$  правил и  $N$  переменных  $x_j$  можно представить в следующем виде:

$$R_1: \text{если } x_1 \in A_1^{(1)}; x_2 \in A_2^{(1)}, \dots, x_n \in A_n^{(1)} \text{ то } y_1 = p_{10} + \sum_{j=1}^N p_{1j} x_j;$$

$$R_M: \text{если } x_1 \in A_1^{(M)}; x_2 \in A_2^{(M)}, \dots, x_n \in A_n^{(M)} \text{ то } y_M = p_{M0} + \sum_{j=1}^N p_{Mj} x_j;$$

где  $A_j^{(k)}$  – значение лингвистической переменной  $x_j$  для правила  $R_k$  с функцией принадлежности (ФП)

$$\mu_A^{(k)}(x_j) = \frac{1}{1 + \left( \frac{x_j - c_j^{(k)}}{\sigma_j^{(k)}} \right)^{2b_j^{(k)}}} \quad (1)$$

$$i = \overline{1, N}; k = \overline{1, M}.$$

В нечеткой сети TSK пересечение условий правила  $R_k$  определяется ФП в форме произведения, то есть

$$\mu_A^{(k)}(x) = \prod_{j=1}^N \left[ \frac{1}{1 + \left( \frac{x_j - c_j^{(k)}}{\sigma_j^{(k)}} \right)^{2b_j^{(k)}}} \right] \quad (2)$$

При  $M$  правилах вывода композиция выходных результатов сети определяется по следующей формуле:

$$y(x) = \frac{\sum_{k=1}^M w_k y_k(x)}{\sum_{k=1}^M w_k} \quad (3)$$

где  $y_k(x) = p_{k0} + \sum_{j=1}^N p_{kj} x_j$ . Присутствующие в этом выражении веса  $w_k$  интерпретируются как степень

выполнения условий правила:  $w_k = \mu_A^{(k)}(x)$ , которые задаются формулой (2).

Нечёткая сеть TSK, которая реализует вывод (3), задается многослойной структурой сети, представленной на рис. 1. В такой сети выделяют 5 слоев.

1. Первый слой выполняет раздельную фаззификацию каждой переменной  $x_j, i = 1, 2, \dots, N$ , определяя для каждого  $k$ -го правила вывода значения функций принадлежности  $\mu_A^{(k)}(x_j)$ , соответственно, к функции фаззификации, которая применяется. Этот параметрический слой с параметрами  $c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$ , которые подлежат адаптации в процессе обучения.

2. Второй слой выполняет агрегирование отдельных переменных  $x_j$ , определяя результирующую степень принадлежности  $w_k = \mu_A^{(k)}(x)$  для вектора  $x$  условиям  $k$ -го правила. Это не параметрический слой.

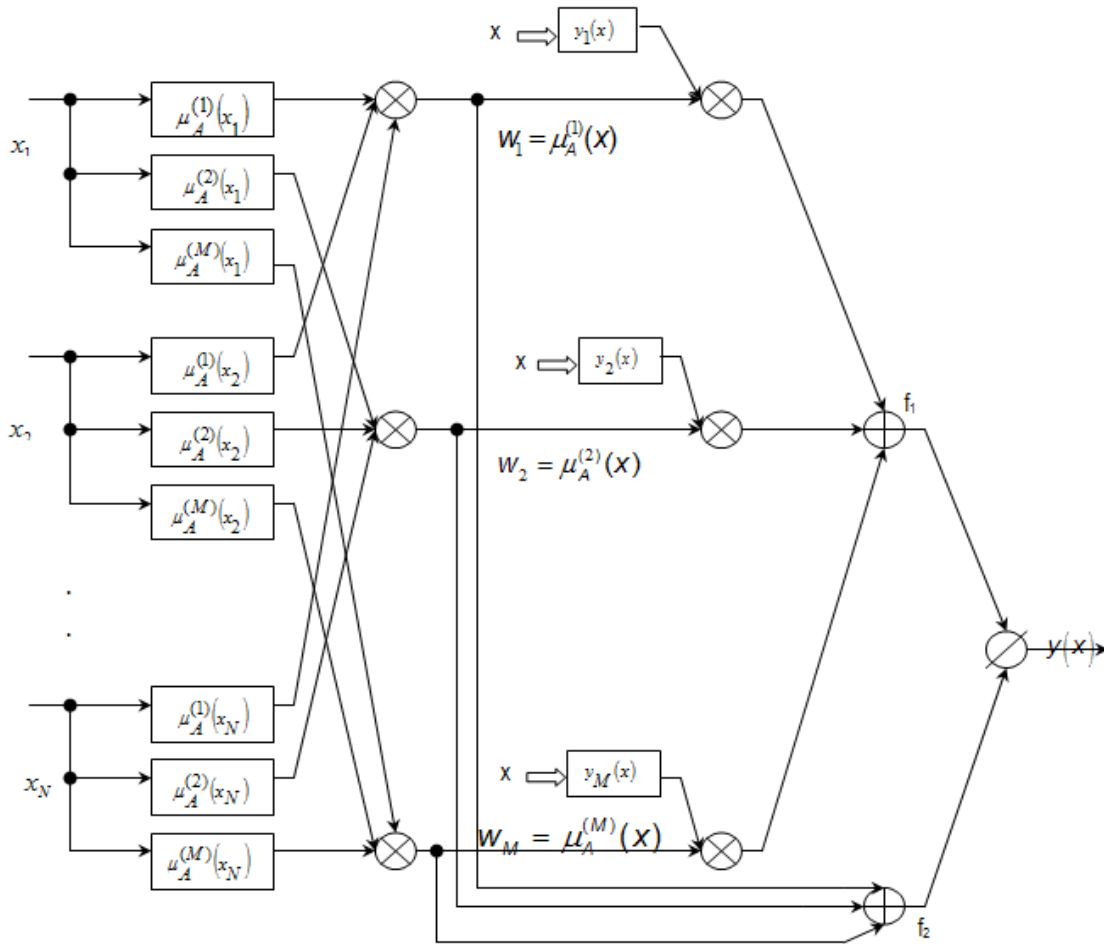


Рис. 1. Структура нечеткой нейронной сети TSK

3. Третий слой представляет собой генератор функций TSK, в котором рассчитывается значение  $y_k(x) = p_{k0} + \sum_{j=1}^N p_{kj} x_j$ . В этом слое происходит также произведение функций  $y_k(x)$  на  $w_k$ , сформированных на предыдущем слое. Это параметрический слой, в котором адаптации подлежат линейные параметры (веса)  $p_{k0}, p_{kj}$  для  $j = \overline{1, N}, k = \overline{1, M}$ , которые определяют функции следствий правил.
4. Четвертый слой составляют 2 нейрона-сумматора, один из которых рассчитывает взвешенную сумму сигналов  $y_k(x)$ , а второй определяет сумму весов  $\sum_{k=1}^M w_k$ .
5. Последний, пятый слой, состоит из единого выходного нейрона. В нем веса поддаются нормализации и вычисляется выходной сигнал  $y(x)$  соответствующий выражению:

$$y(x) = \frac{f_1}{f_2} = \frac{\sum_{k=1}^M w_k y_k(x)}{\sum_{k=1}^M w_k} \quad (4)$$

Это тоже не параметрический слой.

Из приведенного описания следует, что нечеткая нейронная сеть TSK содержит только два параметрических слоя (первый и третий), параметры которых уточняются в процессе обучения. Параметры первого слоя  $\{c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}\}$  будем называть нелинейными, а параметры третьего слоя  $\{p_{kj}\}$  – линейными весами.

Общее выражение для функциональной зависимости (4) для сети TSK задается так:

$$y(x) = \frac{1}{\sum_{k=1}^M \sum_{j=1}^N \mu_A^{(k)}(x_j)} \sum_{k=1}^M \left( p_{k0} + \sum_{j=1}^N p_{kj} x_j \right) \prod_{j=1}^N \mu_A^{(k)}(x_j) \quad (5)$$

Если принять, что в конкретный момент времени параметры условий фиксированы, то функция  $y(x)$  является линейной относительно переменной  $x_j$ .

При наличии  $N$  входных переменных каждое правило  $R_k$  формирует  $(N+1)$  переменную  $p_j^{(k)}$  линейной зависимости  $y_k(x)$ . При  $M$  правилах вывода имеем  $M(N+1)$  линейных параметров сети. В свою очередь каждая функция принадлежности использует три параметра  $(c, \sigma, b)$ , которые подлежат адаптации. При  $M$  правилах вывода получаем  $3MN$  нелинейных параметров. В сумме это дает  $M(4N+1)$  линейных и нелинейных параметров, значения которых должны определяться в процессе обучения. Это очень большая величина. С целью сокращения числа параметров, которые подлежат адаптации, оперируют с меньшим числом независимых ФП. В частности, можно принять, что часть параметров ФП одной переменной  $\mu_A^{(k)}(x_j)$  фиксируется, например,  $\sigma_j^{(k)}$  и/или  $b_j^{(k)}$ .

### Гибридный алгоритм обучения ННС TSK

В [Зайченко, 2008] описан гибридный алгоритм обучения, согласно которому параметры, которые подлежат адаптации, разделяют на две группы. Первая из них состоит из линейных параметров  $p_{kj}$  третьего слоя, а вторая группа – из параметров нелинейной ФП первого слоя. Уточнение параметров проводится в два этапа.

На первом этапе при фиксации отдельных значений параметров ФП (в первом цикле – это значения, которые получены путем инициализации), решая систему линейных уравнений, рассчитывая линейные параметры  $p_{kj}$  полинома TSK. При известных значениях ФП зависимость для выхода можно представить в виде линейной формы относительно параметров  $p_{kj}$

$$y_k(x) = \sum_{k=1}^M w'_k \left( p_{k0} + \sum_{j=1}^N p_{kj} x_j \right) \quad (6)$$

где

$$w'_k = \frac{\prod_{j=1}^N \mu_A^{(k)}(x_j)}{\sum_{r=1}^M \prod_{j=1}^N \mu_A^{(r)}(x_j)}, k = \overline{1, M}. \quad (7)$$

При размерности обучающей выборки  $L(x^{(l)}, d^{(l)})$ ,  $(l=1, 2, \dots, L)$  и замене выходного сигнала сети ожидаемым значением  $d^{(l)}$  получаем систему из  $L$  линейных уравнений вида:

$$\begin{bmatrix} w'_{11} & w'_{11}x_1^{(1)} & w'_{11}x_N^{(1)} & w'_{1M} & w'_{1M}x_1^{(1)} & w'_{1M}x_N^{(1)} \\ w'_{21} & w'_{21}x_1^{(2)} & w'_{21}x_N^{(2)} & w'_{2M} & w'_{2M}x_1^{(2)} & w'_{2M}x_N^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w'_{L1} & w'_{L1}x_1^{(L)} & w'_{L1}x_N^{(L)} & w'_{LM} & w'_{LM}x_1^{(L)} & w'_{LM}x_N^{(L)} \end{bmatrix} \times \begin{bmatrix} p_{10} \\ p_{11} \\ \dots \\ p_{1N} \\ \dots \\ p_{M0} \\ p_{M1} \\ \dots \\ p_{MN} \end{bmatrix} = \begin{bmatrix} d^{(1)} \\ d^{(2)} \\ \dots \\ d^{(L)} \end{bmatrix} \quad (8)$$

где  $w'_{\ell i}$  означает уровень активации (вес) условия  $i$ -го правила при представлении  $\ell$ -го входного вектора  $x^\ell$ . Это выражение можно записать в матричном виде:  $Ap = d$ .

Размерность матрицы  $A$  равна  $L(N+1)M$ . При этом количество строк  $L$  обычно бывает значительно большим, чем количество столбцов  $(N+1)M$ . Решение этой системы уравнений можно получить как обычным методом, так и за один шаг, используя псевдоинверсную матрицу  $A$ :  $p = A^+d$ , где  $A^+$  - псевдоинверсная матрица.

На втором этапе после фиксации линейных параметров  $p_{kj}$  рассчитываются фактические выходные сигналы  $y^{(\ell)}$ ,  $\ell = 1, 2, \dots, L$ , для чего используется линейная зависимость  $y^{(L)} = Ap$ .

После этого рассчитывается вектор ошибки  $\varepsilon = (y - d)$  и критерий

$$E = \frac{1}{2} \sum_{\ell=1}^L (y(x^{(\ell)}) - d^{(\ell)})^2 \quad (9)$$

Сигналы ошибок направляются через сеть в обратном направлении соответственно методу Back propagation, до первого слоя, где могут быть рассчитаны компоненты вектора градиента целевой функции относительно параметров  $c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$ . После вычисления вектора градиента делается шаг спуска по градиентному методу. Соответствующие формулы обучения (для самого простого метода быстрого спуска) принимают вид:

$$c_j^{(k)}(n+1) = c_j^{(k)}(n) - \eta_c \frac{\partial E(n)}{\partial c_j^{(k)}} \quad (10)$$

$$\sigma_j^{(k)}(n+1) = \sigma_j^{(k)}(n) - \eta_\sigma \frac{\partial E(n)}{\partial \sigma_j^{(k)}} \quad (11)$$

$$b_j^{(k)}(n+1) = b_j^{(k)}(n) - \eta_b \frac{\partial E(n)}{\partial b_j^{(k)}} \quad (12)$$

где  $n$  – номер итерации.

После уточнения нелинейных параметров снова запускается процесс адаптации линейных параметров функции TSK (первый этап) и нелинейных параметров (второй этап). Этот цикл повторяется до тех пор, пока не стабилизируются все параметры процесса.

Формулы (10)-(12) требуют расчета градиента целевой функции относительно параметров ФП. Окончательный вид этих формул зависит от вида ФП. Соответствующие формулы градиента целевой функции для одной пары данных  $(x, d)$  принимают вид:

$$\frac{\partial E}{\partial c_j^{(k)}} = (y(x) - d) \sum_{r=1}^M \left( p_{r0} + \sum_{j=1}^N p_{rj} x_j \right) \cdot \frac{\partial w_r'}{\partial c_j^{(k)}} \quad (14)$$

$$\frac{\partial E}{\partial \sigma_j^{(k)}} = (y(x) - d) \sum_{r=1}^M \left( p_{r0} + \sum_{j=1}^N p_{rj} x_j \right) \cdot \frac{\partial w_r'}{\partial \sigma_j^{(k)}} \quad (15)$$

$$\frac{\partial E}{\partial b_j^{(k)}} = (y(x) - d) \sum_{r=1}^M \left( p_{r0} + \sum_{j=1}^N p_{rj} x_j \right) \cdot \frac{\partial w_r'}{\partial b_j^{(k)}} \quad (16)$$

### Поиск оптимальной структуры ННС TSK

В ННС TSK, как и в любой другой НС, используют некоторую выборку данных для обучения и получения прогноза. Но из теоремы неполноты Геделя следует, что выборка всегда неполна. Способом преодоления неполноты выборки данных является принцип внешнего дополнения. В качестве внешнего дополнения используется проверочная выборка, точки которой не участвовали в обучении НС.

Основываясь на сказанном выше, авторами предлагается следующий способ синтезирования ННС TSK оптимальной структуры (т.е. поиск оптимального количества используемых в НС правил):

1. Вся выборка делится на обучающую и проверочную выборки:  $N_{выб} = N_{обуч} + N_{пров}$ .
2. На обучающей выборке  $N_{обуч}$  синтезируется нейронная сеть.
3. На проверочной выборке  $N_{пров}$  по определенному критерию отбирается НС с наилучшей структурой.

### Модификация FOTSK

Модификация ННС TSK под названием FOTSK (Fuzzy Output TSK) разработана авторами для получения не только точечных прогнозов, но и доверительного интервала, который характеризует точность прогноза. ННС FOTSK имеет ту же структуру, что и ННС TSK, только на третьем слое в качестве линейных параметров  $p_{k0}, p_{kj}$  для  $j = \overline{1, N}, k = \overline{1, M}$ , которые определяют функции следствий правил, выступают нечеткие числа треугольного вида, описываемые тройкой параметров  $(\alpha, c_1, c_2)$ , где  $\alpha$  – центр интервала;  $c_1$  – расстояние от центра до нижней границы;  $c_2$  – расстояние от центра до верхней границы;  $c_1, c_2 \geq 0$ .

В гибридном алгоритме обучения НС изменения затронули его первый этап – расчет линейных параметров. Так как  $p_{kj}$  нечеткое треугольное число, то (8) теперь будет иметь следующий вид:

$$\begin{bmatrix} w'_{11} & w'_{11}x_1^{(1)} & w'_{11}x_N^{(1)} & w'_{1M} & w'_{1M}x_1^{(1)} & w'_{1M}x_N^{(1)} \\ w'_{21} & w'_{21}x_1^{(2)} & w'_{21}x_N^{(2)} & w'_{2M} & w'_{2M}x_1^{(2)} & w'_{2M}x_N^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w'_{L1} & w'_{L1}x_1^{(L)} & w'_{L1}x_N^{(L)} & w'_{LM} & w'_{LM}x_1^{(L)} & w'_{LM}x_N^{(L)} \end{bmatrix} \times \begin{bmatrix} \alpha_{10} \\ \alpha_{11} \\ \dots \\ \alpha_{1N} \\ \dots \\ \alpha_{M0} \\ \alpha_{M1} \\ \dots \\ \alpha_{MN} \end{bmatrix} = \begin{bmatrix} d^{(1)} \\ d^{(2)} \\ \dots \\ d^{(L)} \end{bmatrix} \quad (21)$$

или в матричном виде:  $A\alpha = d$ .

Чтобы построенная модель была адекватной, необходимо, чтобы действительное значение выходного сигнала  $d$  принадлежало интервалу неопределенности. Это можно записать следующими неравенствами:

$$\begin{cases} A\alpha - |A|c_1 \leq d \\ A\alpha + |A|c_2 \geq d \end{cases} \quad (22)$$

Для нахождения вектора линейных параметров  $p$  необходимо найти такие значения  $(\alpha, c_1, c_2)$ , которые бы удовлетворяли условиям (22),  $c_1, c_2 \geq 0$ , а суммарная ширина оценочного интервала была бы минимальна. Данные требования сводятся к следующей задаче линейного программирования:

$$\min \left( \text{sum}([|A| \quad |A|]) \times \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right) \quad (23)$$

при условиях:

$$\begin{bmatrix} A & -|A| & A_0 \\ -A & A_0 & -|A| \end{bmatrix} \times \begin{bmatrix} \alpha \\ c_1 \\ c_2 \end{bmatrix} \leq \begin{bmatrix} d \\ -d \end{bmatrix} \quad (24)$$

$$c_1, c_2 \geq 0 \quad (25)$$

где  $A_0$  – матрица идентичной размерности с матрицей  $A$ , состоящая из нулевых элементов; sum – оператор переводящий матрицу в вектор, состоящий из сумм всех элементов строк каждого столбца исходной матрицы.

### Экспериментальные исследования FOTSK в задачах прогнозирования

Для сравнительного анализа исследуемого метода прогнозирования с ННС TSK и с НМГУА были использованы данные с сайта Google Finance по курсу акций компаний IBM, Google, Apple, Coca-Cola и McDonald's, взятые за период с 27.05.2009 по 27.05.2010. Обучающая выборка бралась в размере 0.7 от общей выборки. Прогнозирование производилось на одну эпоху вперед по данным за текущую эпоху.

Полученные прогнозные значения доходностей акций в процентах на 28.05.2010 приведены в табл. 1.

Таблица 1. Результаты применения ННС FOTSK, ННС TSK и НМГУА для решения задач прогнозирования курсов акций.

Метод	Нижняя граница	Центр	Верхняя граница	Реальное значение	Отклонение	СКО внутр.	СКО внешн.	Сложность модели
<b>IBM:</b>								
FOTSK	-3.8064	-1.0902	1.6274	-0.8941	0.1961	1.5100	1.8461	2
TSK	—	-0.2766	—		0.6175	1.2745	1.4415	2
НМГУА	-4.7376	-1.5988	1.4976		0.7047	1.6457	2.1105	1
<b>GOOG:</b>								
FOTSK	-2.3696	0.8376	4.0192	-0.9848	1.8224	6.7583	10.4567	2
TSK	—	0.3489	—		1.3337	6.1955	10.1617	2
НМГУА	-3.2555	0.4508	4.1554		1.4356	7.3152	10.3253	1
<b>AAPL:</b>								
FOTSK	-6.3546	-2.9473	0.4882	1.3933	4.3406	3.2696	6.6176	3
TSK	—	-0.9410	—		2.3344	2.8006	5.0065	3
НМГУА	-0.4310	-0.2806	11.1851		1.6739	3.2572	8.6604	1
<b>KO:</b>								
FOTSK	-1.3660	0.9485	3.2631	0.3318	0.6166	0.6840	0.7755	2
TSK	—	0.3098	—		0.0221	0.5190	0.5889	2
НМГУА	-2.4786	0.1325	2.3320		0.1993	0.5917	0.6981	2
<b>MCD:</b>								
FOTSK	-2.8613	0.3110	3.3783	-0.4911	0.8021	0.9776	1.0751	2
TSK	—	-0.7862	—		0.2951	0.5719	0.8129	3
НМГУА	-1.3899	1.5207	4.4363		2.0118	0.8073	1.2498	3
<b>Суммарно:</b>								
FOTSK	Ширина оценочного интервала: 29.5342				7.7779	13.1994	20.771	
TSK	—				4.6027	11.3615	18.0114	
НМГУА	Ширина оценочного интервала: 35.8990				6.0253	13.6170	23.044	

Как видно из табл. 1, и ННС FOTSK, и НМГУА сошлись в направлении движения курсов акций, правда правильными оказались лишь направления для двух видов ЦБ – IBM и KO. ННС TSK выдала отличное от ННС FOTSK и НМГУА направление прогноза для акций MCD и оказалась права, в сумме правильно определив направление движения курсов для трех ценных бумаг с пяти.

Внешние СКО ННС FOTSK и НМГУА практически одинаковы (в среднем немного ниже у FOTSK). То же справедливо и по отношению к внутреннему СКО. А вот отклонение от реального значения курса у НМГУА получилось меньше. Лидером по трем показателям оказалась ННС TSK, с большим отрывом обогнавшая конкурентов по отклонению от реального значения курса акций. Но стоит отметить, что все пять точечных прогнозов, сделанных на ННС TSK, попали в соответствующие доверительные интервалы построенные ННС FOTSK и НМГУА.

Ширина доверительного интервала у ННС FOTSK на 17.7 % меньше, чем у НМГУА.



Наибольшую ожидаемую доходность ННС FOTSK отдала акциям компании Coca-Cola и при этом правильно определила направление. НМГУА спрогнозировал, что наиболее доходными являются акции компании McDonald's, но направление определил не правильно, а значит, в случае инвестирования в данный вид ЦБ инвестор понесет большие потери. ННС TSK также прогадала, спрогнозировав наибольшую доходность акциям компании Google, которые в действительности показали отрицательное значение прироста доходности, когда наступил момент, на который производился прогноз (28.05.2010).

---

### Заключение

Авторами была разработана ННС FOTSK, дающая не только точечный прогноз, но еще и оценочный интервал к нему. Также, был предложен простой, но эффективный алгоритм, позволяющий определить оптимальное количество используемых правил в ННС TSK/FOTSK для конкретной выборки данных.

С помощью программной реализации ННС FOTSK, ННС TSK и НМГУА были получены прогнозы курсов акций для пяти компаний. По результатам данных экспериментов можно сделать следующие выводы:

1. Представленная модификация хорошо показала себя в решении задач прогнозирования.
2. По величине СКО (как внутреннего, так и внешнего), а также по ширине оценочного интервала ННС FOTSK показала себя лучше НМГУА, но уступила ННС TSK.
3. По суммарному отклонению прогнозных значений от реальных ННС FOTSK уступила и ННС TSK, и НМГУА. Но при этом стоит отметить, что в «акциях-фаворитах» ННС FOTSK, в отличие от ННС TSK и НМГУА, правильно определила направление, что имеет огромное значение при инвестировании.
4. Как показали исследования, большое значение в задачах прогнозирования, особенно курсов ЦБ, играет правильное определение направления. В связи с этим авторы считают, что в дальнейшем при разработке улучшений и модификаций алгоритмов прогнозирования необходимо уделить данному моменту максимальное внимание.
5. При сравнении качества прогнозов от разных методов нужно учитывать не только расхождение с реальными данными, но и то, правильно ли определено направление движения курса акций. Последнее, на взгляд авторов, – приоритетней.

---

### Литература

[Зайченко, 2008] Ю.П. Зайченко. Нечеткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. – К.: «Издательский дом «Слово», 2008. – 344с.

---

### Информация об авторах



**Богдан Гривко** – студент 6 курса НТУУ «КПИ», факультет «Институт Прикладного Системного Анализа», 02225, ул. Каштановая, 13, кв. 102, Киев, Украина;

e-mail: bogdan.grivko@gmail.com

Основная область научных исследований: нейронные сети, искусственный интеллект, портфельная теория



**Юлия Проскурня** – студентка 6 курса НТУУ «КПИ», факультет «Институт Прикладного Системного Анализа», 02225, ул. Каштановая, 13, кв. 102, Киев, Украина;

e-mail: julia.proskurnya@gmail.com

Основная область научных исследований: нейронные сети, искусственный интеллект, портфельная теория

## ПРОГНОЗИРОВАНИЕ В ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИИ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Галина Сетлак

**Аннотация:** В данной работе исследуются возможности использования для решения задач прогнозирования искусственных нейронных сетей. Предлагается система построения прогнозов с использованием нескольких методов прогнозирования, качество и точность которых оценивается экспертной системой. Сравнение полученных прогнозов с использованием нейросетевых моделей производится на основе решения задач при помощи традиционных методов прогнозирования.

**Ключевые слова:** прогнозирование, искусственные нейронные сети, экспертная система.

**ACM Classification Keywords:** H. Information Systems, H.4 INFORMATION SYSTEMS APPLICATIONS, H.4.2 Types of Systems - Decision support. : I. Computing Methodologies, [1.2 Artificial Intelligence](#), J. Computer Applications

**Conference topic:** Pattern Recognition and Forecasting

---

### Введение

Прогнозирование – одна из самых необходимых, но при этом и самых сложных, задач стратегического анализа. Проблемы при ее решении обусловлены многими причинами – недостаточное качество и количество исходных данных, изменения среды, в которой протекает процесс, воздействие субъективных факторов. Но именно качественный прогноз является ключом к решению таких задач в управлении предприятием как оптимизация финансовых потоков, бюджетирование, оценка инвестиционных проектов и многих других.

Прогнозирование упрощённо можно определить как предсказание будущих событий. Согласно терминологии научной дисциплины прогностики, прогноз – это научно обоснованное суждение о возможных состояниях объекта в будущем и(или) об альтернативных путях и сроках их осуществления [Коваленко И.И., Бидюк П.И., Баклан И.В., 2001]. Целью прогнозирования является уменьшение риска при принятии решений. Ошибка прогноза зависит от используемого метода прогнозирования и используемых для его подготовки данных.

Раньше для решения задач прогнозирования традиционно использовались такие методы математической статистики, как регрессионный, факторный анализ или метод функции тренда и другие.

В настоящей работе исследуются возможности использования для решения задач прогнозирования искусственных нейронных сетей. Предлагается система построения прогнозов с использованием нескольких методов прогнозирования, качество и точность которых оценивается экспертной системой. Сравнение полученных прогнозов с использованием нейросетевых моделей производится на основе решения задач при помощи традиционных методов прогнозирования: в зависимости от решаемой задачи используются метод скользящего среднего, авторегрессии, экспоненциального сглаживания, или функции тренда и тенденции развития. В результате оценки системой полученных прогнозов выбирается наилучший (оптимальный) или строится комплексный прогноз, который является комбинацией прогнозов, построенных при помощи различных методов.

Целью настоящих исследований является разработка метода построения интеллектуальной системы выбора оптимального прогноза спроса и продажи изделий, изготавливаемых предприятием, функционирующем в условиях рыночной экономики.

В хозяйственной деятельности наиболее важные для предприятия решения принимаются на основе выполнения целого ряда прогнозов. Основные прогнозы, необходимые в процессе принятия решений в предприятиях можно описать в следующей зависимости:

- Для разработки главных целей функционирования организации и долгосрочных бизнес-планов, а также выбора стратегии развития предприятия необходимы прогнозы: спроса на товары и положения на рынках, прогнозы общих экономических условий (инфляция, ситуация в отрасли), прогнозы продажи и цен сырья, прогнозы внутренних условий в предприятии (производственных затрат, производства).

- Для планирования разработки и введения нового продукта на рынок требуется долгосрочный прогноз того, каким спросом он будет пользоваться, каковы предполагаемые цены. Отдел маркетинга нуждается в прогнозах, объясняющих ситуацию на рынке: прогноз глобального спроса по сегментам рынка и изделиям, а также прогноз продажи.

- Для разработки производственных планов по величине и ассортименту производства необходимы прогнозы продаж, прогнозы производственных средств (сырья и материалов, энергии и т.д.). В процессе принятия всех решений на предприятиях прогнозируются затраты на хозяйственную деятельность на основе прогнозы предполагаемой прибыли, финансовых результатов.

Между всеми выше представленными прогнозируемыми данными имеются определённые зависимости, которые приводят к необходимости соблюдения определённой очерёдности прогнозирования и, безусловно, выделяются наиболее важные прогнозы, от точности которых зависит успех предприятия: прогнозы спроса и продажи. Прогноз продажи играет исключительно важную роль в управлении производством, так как на его основе определяется ожидаемый доход, т.е. основной источник финансирования деятельности предприятия. Прогнозные величины продажи являются исходными данными в процессе стратегического и оперативного планирования. Поскольку при решении реальных задач прогнозирования спроса на продукты возникают проблемы с получением и подготовкой входных данных, в данной работе решается обобщённая задача прогнозирования объёма продаж с учётом фактора покупательского спроса.

---

### Постановка задачи

---

Задача прогнозирования в общем случае сводится к получению оценки будущих значений упорядоченных во времени данных на основе анализа уже имеющихся ретроспективных данных. Прежде всего, необходимо решить два вопроса: что является прогнозируемой величиной, и что является входными данными.

В большинстве случаев прогнозируемой величиной являются значения временного ряда на интервале  $[T(n+1), T(n+f)]$ , где  $T(n)$  – текущий момент времени, а  $f$  – интервал прогнозирования. Иногда возникает необходимость предсказать не значения временного ряда на заданном интервале, а вероятность того, что он будет вести себя каким-то образом (возрастать, убывать, находиться в некоторых пределах и т.д.). Форма взаимосвязи прогнозируемого явления с другими явлениями, объектами и процессами может быть представлена в виде регрессионного уравнения типа:

$$y = f(x_1, x_2, \dots, x_m).$$

В данных исследованиях задача прогнозирования объёма продаж решается на основе ежеквартальных данных о динамике реализации продукции в период времени от 2006 по 2009 год (предприятием изготавливающим приборы домашнего хозяйства, в том числе пылесосы), а также данных, полученных в результате подготовленного и выполненного консалтинговой фирмой анкетного опроса.

В качестве исходных данных выбираем из имеющихся данных максимальное число значащих факторов. Это означает выбор интервала наблюдения, то есть, по какому количеству предшествующих значений временного ряда осуществляется прогноз и определение дополнительных факторов, влияющих на поведение прогнозируемой величины, называемых также экзогенными (внешними) факторами. В работе учитываются такие дополнительные факторы, как рекламная компания, конъюнктура рынка и экспансия конкуренции. Ретроспективные ежеквартальные данные о динамике реализации продукции и положения на рынке в период времени от 2006 по 2009 год представлены в таблице 1.

Таблица 1. Ретроспективные ежеквартальные данные о динамике реализации продукции в период времени от 2006 по 2009 год и положения на рынке

Период времени	Объём продажи (в злотых)	Затраты на рекламу (в злотых)	Спрос на изделия	Экспансия конкуренции
1_2006	86158	8000	средний	небольшая
2_2006	89134	6000	большой	небольшая
3_2006	72250	4000	низкий	небольшая
4_2006	89912	0	большой	небольшая
1_2007	82250	6000	средний	средняя
2_2007	86159	8000	средний	средняя
3_2007	69925	5000	низкий	средняя
4_2007	78390	5000	большой	средняя
1_2008	65930	4000	низкий	выше средней,
2_2008	73192	4000	средний	выше средней,
3_2008	61910	2000	низкий	выше средней,
4_2008	69668	2000	большой	выше средней,
1_2009	69191	0	Средний	выше средней,
2_2009	65498	0	средний	большая
3_2009	59267	0	низкий	большая
4_2009	64057	3000	низкий	большая

Спрос на изделия рассматривается как обобщенное положение и конъюнктура на рынке и описывается одним из возможных значений: (низкий, средний, большой). Экспансия конкуренции описывается величинами из набора: (небольшая, средняя, выше средней, большая).

Рассмотрим общую характеристику предлагаемого подхода разработки интеллектуальной системы прогнозирования.

### **Общая характеристика предлагаемого подхода разработки интеллектуальной системы прогнозирования**

Для решения поставленных выше задач и построения моделей нейронных сетей используем программный пакет Statistica Neural Networks и интеллектуальную систему поддержки принятия решений IDSS (Intelligent Decision Support System), разработанную автором и представленную в [Сетлак Г., 2002], [Сетлак Г., 2004], в состав которой входит экспертная система и модуль PROGNOZY.

В программном пакете Statistica Neural Networks (STNN) реализован широкий спектр технологий, которые можно успешно использовать для решения задач прогнозирования и обеспечивающих весь необходимый цикл обработки данных, начиная от извлечения и предобработки данных и кончая построением нейросетевых моделей. В пакете STNN для решения задач прогноза временных рядов можно применять сети нескольких типов: многослойный персептрон (MLP), сеть радиальной базисной функции (RBF), вероятностная нейронная сеть (PNN) и обобщённо-регрессионные нейронные сети (GRNN).

Выбор входных переменных является исключительно важным этапом при построении нейронной сети. С целью устранения из входных данных несущественных и редко встречающиеся факторов необходимо провести предобработку данных: восстановить пропущенные данные, устранить аномальные выбросы, убрать высокочастотные шумы. Качественная предобработка позволяет значительно улучшить качество прогноза. Предобработка данных и выбор нужных входных переменных при разведочном анализе данных выполняется при помощи модуля Нейро-генетического отбора входных данных (Neuro-Genetic Input Selection), входящего в программный пакет STNN. Генетические алгоритмы хорошо приспособлены для задач такого типа, поскольку они позволяют производить поиск среди большого числа комбинаций при наличии внутренних зависимостей в переменных.

Строится нейросетевая модель на основе данных о продаже в оптовых ценах за 2006-2009 год (представленные в первой и второй колонках в таблице 1. согласно со следующим алгоритмом:

1) Создание и оптимизация выборки. Выборка создаётся на основе ежеквартальных данных за 2006-2009 г.г, что позволяет учитывать цикличность сезонного спроса. Предобработка данных (устранение незначущих факторов, восстановление пропущенных значений, устранение выбросов, фильтрация высокочастотных шумов) и выбор нужных входных переменных выполняется при помощи модуля Нейро-генетического отбора входных данных (Neuro-Genetic Input Selection), входящего в программный пакет STNN.

2) Следующий шаг – это построение нейросетевой модели. На этом шаге необходимо решить ряд специфических подзадач: выбор структуры нейронной сети, алгоритма обучения и другие. При построении сети для прогноза временного ряда используем входную переменную – объём продаж, в качестве выходной переменной (её будущие значения) и задаём ей тип Input/Output.

Для поиска лучшей архитектуры и выбора оптимального метода обучения был использован модуль «Интеллектуальный исследователь задач» (Intelligent Problem Solver - IPS), который содержит программный пакет STNN. Модуль IPS - „Интеллектуальный исследователь задач” используется на этапе конструирования нейронной сети, позволяет автоматически выбрать тип сети, архитектуру и способ обучения, рассматривая поиск лучшего варианта как оптимизационную задачу. С его помощью исходные данные разбиваются случайным образом на три подмножества: 18 позиций составили обучающую выборку, 9 позиций — тестовое множество и 9 позиций — контрольное.

Затем производится поиск нескольких лучших сетей, среди которых были отобраны 10 разного качества, определяемого средней ошибкой, и сложности для каждого типа нейронных сетей. Эксперименты повторялись многократно, оценивались возможности прогнозирования временных рядов всех выше представленных типов сетей.

3) После построения нейронной сети на следующем этапе выполняется её обучение. В зависимости от типа нейронной сети использовался соответствующий алгоритм её обучения:

Многослойный перцептрон обучался при помощи алгоритма обратного распространения ошибок (Back Propagation) или метода спуска по сопряжённым градиентам (Conjugate Gradient Descent).

Нейронные сети RBF радиальной базисной функции при помощи алгоритмов: K-средних для назначения радиальных центров (K-Means), K-ближайшего соседства для определения отклонений (радиусов) радиальных элементов (K-Nearest Neighbour) и минимально - квадратичного алгоритма оптимизации для линейного выходного слоя (Pseudo - Invert).

IPS выбрал лучшую сеть типа RBF, представленную на рис.1 и с параметрами описанными в таблице 2.

Таблица 2. Результаты проектирования временного ряда продажи при помощи нейронных сетей

Type	Error	Inputs	Hidden	Performance
MLP	194.388	1	7	0.02483
MLP	236.0879	1	4	0.03048
MLP	199.1786	1	8	0.02569
MLP	185.4026	1	13	0.0236
MLP	147.7128	1	8	0.01853
MLP	100.7422	1	8	0.013
MLP	86.27347	1	13	0.01109
MLP	63.88906	1	10	0.008247
MLP	55.54256	1	20	0.007051
RBF	8.769e-12	1	7	8.136e-16
MLP	55.64438	1	13	0.007137
RBF	4.288e-12	1	7	6.538e-16

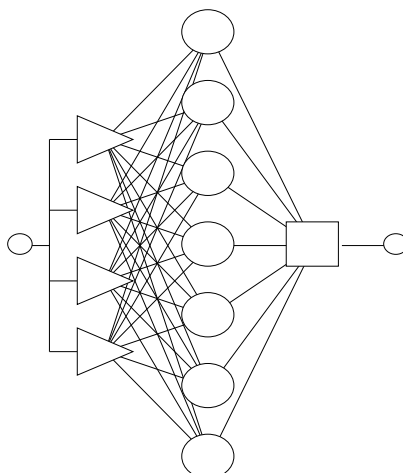


Рис.1. Нейронная сеть типа RBF, выбранная IPS для прогнозирования временного ряда.

Результаты прогнозирования с использованием нейронных сетей представлены на рис.2.

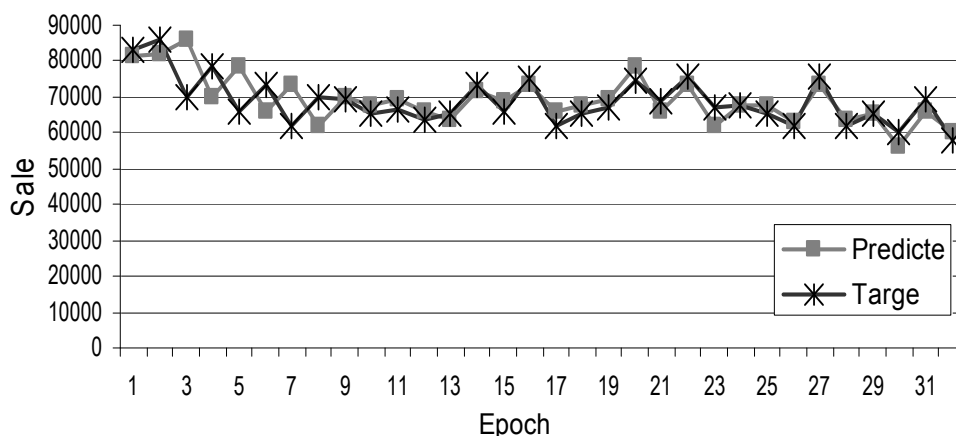


Рис.2. Результаты прогнозирования при помощи нейронных сетей

4) Проверка погрешности и оценка качества принятой прогнозной модели и метода выполняется в исследованиях с использованием программного модуля PROGNOZY, входящего в состав IDSS.

Программный модуль PROGNOZY подготавливает необходимую информацию для разработки прогнозов. В нем создаются заказы на изготавливаемую продукцию, обрабатываются информации, приобретённые на основе анкетных опросов клиентов, которые используются для прогнозирования спроса, выполняется также предварительная обработка и анализ данных, необходимых для решения задач прогнозирования.

Модуль PROGNOZY содержит также необходимые инструменты и процедуры для решения таких задач, как:

предварительная обработка и анализ данных, необходимых для разработки прогнозов;

выбор метода прогнозирования,

оценки качества принятой прогнозной модели и метода.

Все известные показатели, используемые для оценки качества полученного прогноза можно разделить на три группы: абсолютные, сравнительные и качественные. К абсолютным показателям качества прогнозов относятся такие, которые позволяют количественно оценить величину ошибки прогноза в единицах измерения прогнозируемого объекта или в процентах. В практике чаще всего используются следующие ошибки точности [Manikowski A., Tarapaty Z., 2000]:

абсолютная ошибка прогноза, выраженная в процентах и рассчитываемая по формуле (mean absolute percentage error):

$$MAPE = \frac{1}{T^*} \sum_{t=1}^{T^*} \left( \frac{y_t^* - y_t}{y_t} \right) \cdot 100 \quad (1)$$

среднеквадратическая ошибка прогноза, рассчитываемая по формуле:

$$RME = \sqrt{\frac{1}{T^*} \sum_{t=1}^{T^*} (y_t^* - y_t)^2} \quad (2)$$

среднеквадратическая ошибка прогноза, выраженная в процентах и рассчитываемая по формуле (root average squared error):

$$RASE = \sqrt{\frac{1}{T^*} \sum_{t=1}^{T^*} \left( \frac{y_t^* - y_t}{y_t} \right)^2} \cdot 100 \quad (3)$$

В общем случае вопрос о качестве прогноза сводится к допустимой средней ошибке и коэффициенту корреляции. В исследованиях выполнялось сравнение прогнозов полученных при помощи нейронных технологий с прогнозами, полученными с использованием классических статистических методов. Было показано, что в большинстве случаев нейросеть дает более качественный результат, чем модели ARIMA, что показано в таблице 3.

Таблица 3. Сравнение ошибок и качества прогнозов, полученных с использованием нейронных сетей и статистических методов

Реальные значения	Нейронная сеть		ARIMA	
	RASE	RME	RASE	RME
86158	4,12	0,0331	4,9872	0,0491
89134	6,5361	0,07075	16,594	0,17075
72250	6,7921	0,0681	7,9987	0,6391
89912	3,7561	0,03853	11,4891	0,3592
82250	2,9595	0,02496	9,8576	0,09867
86159	6,3107	0,06077	12,9857	0,36857
69925	2,4973	0,02423	9,5637	0,02423
78390	12,2941	0,12324	362,2941	0,35823
69930	0,1389	0,00139	2,8639	0,02875
73192	0,1598	0,00116	0,2297	0,02139
61910	0,2012	0,00192	2,9017	0,02978
63668	2,1134	0,01901	2,1134	0,01901

Выбор конкретного метода прогнозирования осуществляется исходя из нескольких критериев, в том числе с учетом требований, предъявляемых к информационному обеспечению решаемых задач.

## Выводы

В данной работе предложен подход к построению интеллектуальных систем прогнозирования в хозяйственной деятельности предприятия, основанный на использовании искусственных нейронных сетей и оценки качества получаемых прогнозов при помощи экспертной системы.

При помощи построенной интеллектуальной системы на основе нейронных сетей имеем возможность строить краткосрочные и среднесрочные прогнозы. Положительно на качество прогноза влияет дополнительная информация о внешней среде (курс доллара, рекламная поддержка и т.д.).

Одним из важнейших преимуществ предлагаемого подхода прогнозирования является отсутствие необходимости в строгой математической спецификации модели, что особенно ценно при анализе плохо

формализуемых процессов. Это означает, что при наличии достаточно развитых и удобных инструментальных программных средств пользователь может при построении модели прогнозируемого процесса руководствоваться собственным опытом и интуицией. При этом следует отметить, что комплексный учет всех факторов (эндогенных и экзогенных) значительно повышает качество прогноза.

По сравнению с традиционными методами прогнозирования (математической статистики, регрессионного, факторного анализа и т.д.), подход, основанный на применении искусственных нейронных сетей, обладает следующими преимуществами:

- Во-первых, в результате построения нейросетевой модели создается единая для всех задач вычислительная парадигма. Используя нейронные сети со сравнительно небольшим числом нейронов, можно решать достаточно сложные задачи прогноза. А также, используя одну построенную нейронную сеть, можно решать одновременно несколько задач прогноза.

- Поскольку нейронные сети представляют собой обучающиеся модели, при поступлении новых данных достаточно "доучивать" уже имеющуюся нейронную сеть, либо "переучивать" ее для обработки данных из другого региона или другой фирмы.

- В нейронных сетях можно использовать любое количество независимых и зависимых признаков.

- В нейронной сети имеется процедура подсчета и оценки значимости независимых признаков и возможность минимизации их числа.

- Качество получаемых прогнозов оценивается при помощи экспертной системы.

Анализ результатов показал, что получаемые прогнозы имеют достаточно высокий уровень точности, что позволит использовать их с целью снижения рисков при принятии решений, основанных на нейросетевых исследованиях.

Принимая во внимание надежность функционирования нейронных сетей и возможности разработки базы знаний экспертной системы для оценки качества прогнозов и выбора оптимального метода в будущем, можно предположить, что интеллектуальную систему прогнозирования можно будет с успехом применять в прогнозировании экономических процессов, для которых классические методы бессильны.

Дальнейшие перспективы развития разработанного подхода и программного продукта связаны с повышением чувствительности настраиваемых параметров нейронной сети и расширением модуля PROGNOZY, а также базы знаний экспертной системы с целью совершенствования системы оценки качества прогнозов, что в конечном счете приведет к уменьшению ошибки прогноза, а также к еще большей его универсализации.

---

## Литература

[Коваленко И.И., Бидюк П.И., Баклан И.В., 2001] Системный анализ и информационные технологии в управлении проектами, Киев, Экономика и право, 2001, 267 с..

[Manikowski A., Tarapaty Z., 2000] Manikowski A., Tarapaty Z.: Prognozowanie i symulacja rozwoju przedsiębiorstw, Wyd. Wyższej Szkoły Ekonomicznej, Warszawa, 2000, 334 str.

[Сетлак Г., 2002] Сетлак Г. Интеллектуальная система поддержки принятия решений в нечёткой среде, Искусственный интеллект, N3, 2002, стр. 428-438.

[Сетлак Г., 2004] Сетлак Г.: Интеллектуальные системы поддержки принятия решений, Изд. Логос, Киев, 2004, 252 с.

[Statsoft, 1999] Statsoft: Statistica Neural Networks, Addendum for Version 4.0, Statsoft Inc., 1999.

---

## СВЕДЕНИЯ ОБ АВТОРЕ

**Galina Setlak**, D.Sc, Ph.D., Eng., Associate Professor, Rzeszow University of Technology, Department of Computer Science, W. Pola 2 Rzeszow 35-959, Poland, Phone: (48-17)- 86-51-433, gsetlak@prz.edu.pl



## СОЗДАНИЕ ВЫЧИСЛИТЕЛЬНЫХ МОДЕЛЕЙ В ОБЛАСТИ ГОРЕНИЯ КОНДЕНСИРОВАННЫХ СИСТЕМ СРЕДСТВАМИ DATA MINING

**Виктор Абруков, Сергей Абруков, Елена Карлович**

**Аннотация:** Представлены возможности средств *Data Mining*, в частности искусственных нейронных сетей (ИНС) для моделирования и прогнозирования характеристик горения конденсированных систем (КС). Впервые получены вычислительные модели, позволяющие прогнозировать характеристики горения КС. Они позволяют прогнозировать профили температуры в волне горения КС, скорость горения различных по составу КС при различных давлениях и начальных температурах, определять состав КС, обеспечивающий необходимую скорость горения при заданном давлении и т.д. Описаны и проиллюстрированы методы создания таких моделей. Полученные результаты показывают, что ИНС могут рассматриваться как хорошее средство аппроксимации многомерных экспериментальных данных, позволяющее обобщать и прогнозировать связи между переменными эксперимента и теории, как быстрый инженерный калькулятор, специализированный для решения задач исследования процесса горения КС, как средство получения новых «экспериментальных» результатов и выявления новых неизвестных ранее закономерностей горения, как хорошее средство представления и хранения ранее полученных экспериментальных результатов.

**Ключевые слова:** *Data Mining*, искусственные нейронные сети, горение, вычислительные модели,

---

### Введение

К настоящему времени накоплено много экспериментальных данных по характеристикам горения различных конденсированных систем (КС). Вопрос заключается в следующем – можно ли их обобщить и на этой основе представить в виде модели, позволяющей предсказывать закономерности горения ранее не исследованных КС или для ранее не исследованных условий?

Очевидно, что такие параметры КС, условий горения и характеристики горения, как: состав КС, начальная температура образца КС, давление, максимальная температура пламени, распределение температуры в пламени, излучение пламени, скорость горения, связаны между собой. Вопрос – как мы можем обобщить эти связи?

В данной работе представляются результаты применения средств *Data Mining* (часто называемых средствами интеллектуального анализа данных) для создания многофакторных моделей обобщающих закономерности горения, выраженные в экспериментальных данных. Данная задача поставлена впервые, мы не нашли в научной литературе примеров применения *Data Mining* для решения данной задачи.

---

### Методы исследования, результаты и их обсуждение

*Data Mining* - комплекс современных средств предобработки данных, их анализа и визуализации результатов анализа. *Data Mining* включает в себя, в частности, такие средства, как: факторный и корреляционный анализ, «деревья решений», искусственные нейронные сети (ИНС), самоорганизующиеся карты Кохонена. С точки зрения создания многофакторных вычислительных моделей, ИНС, которые могут рассматриваться как универсальные аппроксиматоры многомерных функций, играют главную роль. Остальные средства служили нам для предобработки данных и для их предварительного анализа.

К настоящему времени на основе экспериментальных данных по горению конденсированных систем, взятых из научной литературы, нами получены четыре вычислительные ИНС-модели горения КС.

### 1.1. Модель «Температурный профиль - скорость горения»

Данная модель позволяет определять профиль температуры в волне горения КС через скорость горения,  $U$ , давление,  $p$ , и удельную теплоту сгорания в близповерхностном слое горящего образца,  $q$  (или максимальную температуру пламени,  $T_{\max}$ ).

Схема построения модели была следующая. Были собраны данные по температурным профилям в газовой фазе горения конденсированных систем [Зенин, 1980, Льюис, 1976], измеренным для различных систем, скоростей горения и давлений. Затем была выбрана соответствующая собранной базе данных архитектура ИНС и проведено ее «обучение». Обучение заключалось в том, что различные наборы значений скоростей горения, давления, удельной теплоты сгорания (или максимальной температуры пламени) и координат волны горения подавались на входной слой ИНС, а соответствующие значения температуры устанавливались в выходном слое ИНС и с помощью известного метода обучения ИНС – метода «обратного распространения ошибки» создавалась вычислительная ИНС модель горения.

Эта модель представляет собой модель типа «чёрного ящика», она не имеет аналитического выражения. Полученный «чёрный ящик» может использоваться для определения (предсказания) температурного профиля в газовой фазе горения конденсированных систем примерно такого же типа, для которого была собрана база данных следующим образом. Значение давления и скорости горения (они должны быть измерены) и теплота сгорания или максимальная температура (они могут быть оценены или рассчитаны) устанавливаются во входном слое ИНС. После этого «чёрный ящик» мгновенно вычисляет соответствующий этим значениям профиль температуры.

Примеры таких вычислительных моделей получены нами для смесевых КС на основе перхлората аммония и нитроцеллюлозы. Ниже представлены результаты и их обсуждение.

Таблица 1

Часть данных, использованная для создания ИНС-модели температурного профиля в газовой фазе горения смесевых КС на основе перхлората аммония.

$U$ , см/сек	$p$ , атм	$q$ , кал/г	$x$ , мм	$T$ , °C
0.2	5	110	0	400
0.2	5	110	0.1	1100
0.2	5	110	0.2	1300
0.2	5	110	0.3	1480
...	...	...	...	...
0.2	5	110	0.8	1900
0.2	5	110	0.9	1930
0.2	5	110	1	1940
0.8	100	170	0	600

$U$ , см/сек	$p$ , атм	$q$ , кал/г	$x$ , мм	$T$ , °C
0.8	100	170	0.1	1595
...	...	...	...	...
0.8	100	170	1	2400
0.5	20	210	0	600
...	...	...	...	...
0.5	20	210	0.5	2250
0.8	40	230	0	500
...	...	...	...	...

Результаты, полученные с помощью ИНС-модели, представлены в Таблице 2 и на рис. 1.

Таблица 2

Результаты проверки работы ИНС-модели для смесевых КС на основе перхлората аммония. Условия горения:  $p=40$  атм,  $Q=150$  кал/г,  $U=0.4$  см/сек.  $T$  - температура, измеренная термпарой,  $T'$  - температура, вычисленная ИНС-моделью.

$x$ , мм	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$T$ , °C	600	1150	1550	1650	1710	1800	1950	1980	2020	2100	2200
$T'$ , °C	577	1225	1454	1636	1770	1882	1982	2068	2141	2198	2242

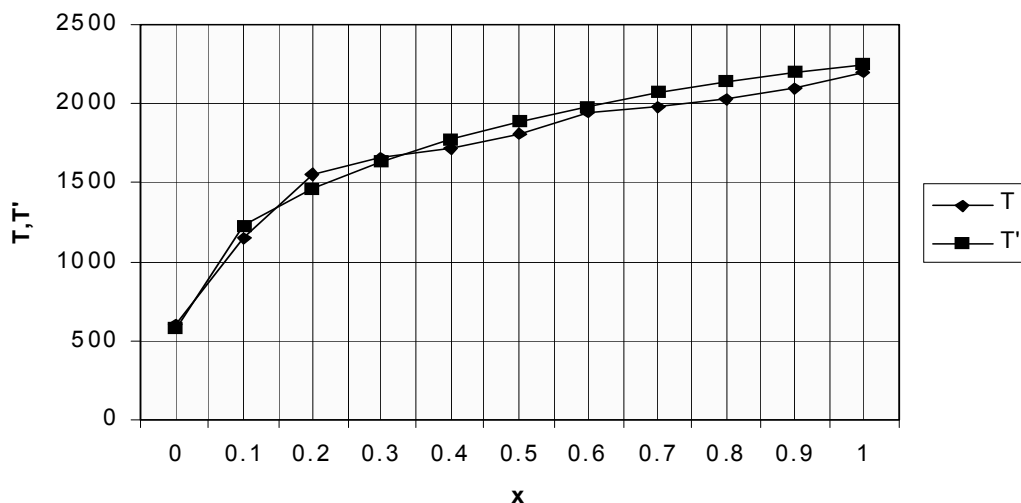


Рис. 1. Результаты проверки ИНС-модели для смесевых КС на основе перхлората аммония. Условия горения:  $p=40$  атм,  $Q=150$  кал/г,  $U=0.4$  см/сек.  $T$  - температура, измеренная термопарой,  $T'$  - температура, вычисленная ИНС-моделью.

Другой пример данных использованных для построения ИНС-модели температурного профиля в газовой фазе горения КС представлен в табл. 3, а результаты проверки – в табл. 4 и на рис.2. Здесь, мы использовали экспериментальные данные по горению нитроцеллюлозы [Льюис, 1976].

Анализ табл. 2, 4 и рис. 1, 2 показывает, что ИНС-модель «Температурный профиль - скорость горения» позволяет предсказывать температурный профиль в газовой фазе горения КС по данным о скорости горения, давления и удельной теплоте сгорания (или максимальной температуре пламени).

По нашему мнению, этот способ получения температурных профилей может быть многообещающим в случаях, когда мы имеем пробелы в экспериментальных результатах для исследуемых КС и хотели бы заполнить их. Мы полагаем, что этот путь определения температурных профилей может рассматриваться в перспективе как уникальный инструмент для случаев, когда экспериментальные данные по профилям температуры полностью отсутствуют, но мы хотели бы иметь их оценку. Чтобы иметь возможность выполнять эту оценку, необходимо собрать экспериментальные данные по профилям температуры для разнообразных КС при различных условиях. После этого можно поставить задачу построения обобщенной ИНС-модели «Температурный профиль - скорость горения» для КС в целом.

Таблица 3: Часть данных, использованная для создания ИНС-модели температурного профиля в газовой фазе горения нитроцеллюлозы.

$p$ , атм	$U$ , мм/сек	$T_{max}$ , °C	$x$ , мм	$T$ , °C
49	25	1500	0	200
49	25	1500	0.1	790
49	25	1500	0.2	1000
49	25	1500	0.4	1270
49	25	1500	0.6	1380
49	25	1500	0.8	1400
49	25	1500	1.0	1430
49	25	1500	1.2	1450
49	25	1500	1.4	1470
42	22	1400	0	300
42	22	1400	0.1	650
42	22	1400	0.2	950

$p$ , атм	$U$ , мм/сек	$T_{max}$ , °C	$x$ , мм	$T$ , °C
42	22	1400	0.4	1230
42	22	1400	0.6	1300
42	22	1400	0.8	1380
42	22	1400	1.0	1393
42	22	1400	1.2	1397
42	22	1400	1.4	1400
20	19	1200	0	270
20	19	1200	0.1	470
20	19	1200	0.2	700
20	19	1200	0.4	1095
20	19	1200	0.6	1200
...	...	...	...	...

Таблица 4

Результаты проверки ИНС-модели для нитроцеллюлозы. Условия горения:  $p=28$  атм,  $U=2,0$  см/сек,  $T_{\max} = 1400$  °С.  $T$  - температура, измеренная термопарой,  $T'$  - температура, вычисленная ИНС-моделью.

x, мм	0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4
T, °С	250	580	850	1190	1250	1270	1300	1370	1380
T', °С	302	506	829	1149	1239	1283	1321	1355	1379

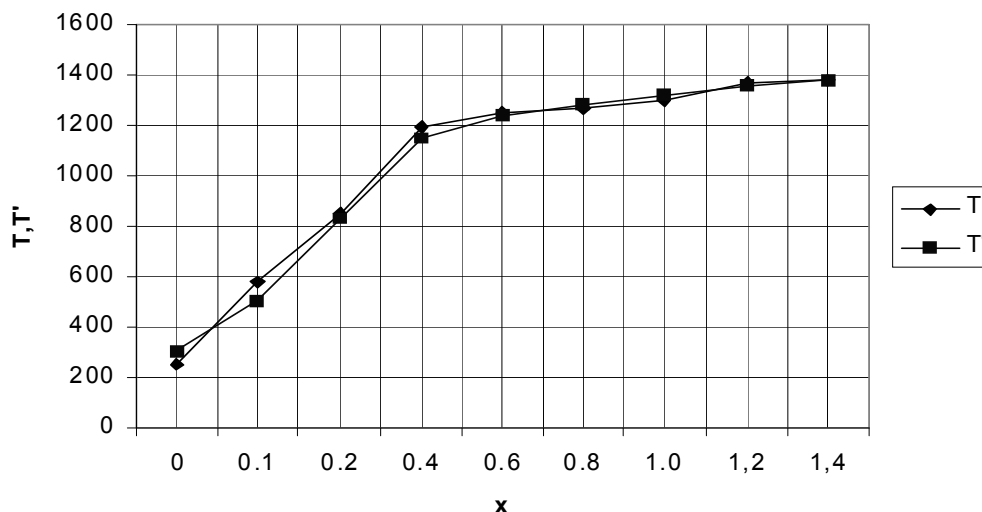


Рис. 3. Результаты проверки ИНС-модели для нитроцеллюлозы. Условия горения:  $p=28$  атм,  $U=2,0$  см/сек,  $T_{\max} = 1400$  °С.  $T$  - температура, измеренная термопарой,  $T'$  - температура, вычисленная ИНС-моделью.

## 1.2. Модель “Скорость горения – давление и состав топлива”

Данная ИНС-модель устанавливает соотношение между значением скорости горения с одной стороны, и количеством нитрата аммония и перхлората аммония, размером частиц перхлората аммония и давлением с другой стороны. Использовались данные работы [Makota, 2008]. Эта модель позволяет предсказывать значение скорости горения конденсированных систем такого типа в различных условиях.

## 1.3. Модель “Состав топлива – давление и скорость горения”

Данная ИНС-модель устанавливает соотношение между количеством перхлората аммония и размером его частиц с одной стороны и значением скорости горения, количеством нитрата аммония и значением давления с другой стороны. Для ее построения также использовались данные работы [Makota, 2008]. Эта модель позволяет предсказывать состав топлива необходимый для получения требуемой скорости горения при заданном давлении.

## 1.4. Модель “Скорость горения – давление, начальная температура, теплота взрыва монотоплива”.

Данная ИНС-модель устанавливает соотношение между значением скорости горения монотоплив способных к самостоятельному горению: гексоген (RDX), октоген (HMX), перхлорат аммония, топливо CL-20, динитрамид аммония (ADN), нитроформат гидразина (HNF) с одной стороны, и давлением, начальной температурой и теплота взрыва монотоплива с другой стороны. Эти вещества являются основными компонентами многих КС. Использовались данные работы [Atwood, 1999]. Эта модель позволяет предсказывать значение скорости горения монотоплив в различных условиях.

## Заключение

Мы полагаем, что аналогичные модели могут быть получены для различных групп КС, а в перспективе и для КС в целом и что средства Data Mining позволят существенно увеличить значение уже полученных

экспериментальных результатов, а также получить новые “экспериментальные” результаты и выявить новые неизвестные ранее закономерности горения.

Помимо чисто научных задач, это позволит облегчить решение и практически важных задач, например, задачи подбора состава топлива и поиска новых топлив для обеспечения требуемых скоростей горения и других характеристик. В перспективе, обобщенную модель горения КС в целом можно будет рассматривать как удобную в использовании базу знаний (Knowledge Data Base) закономерностей горения конденсированных систем. Часть ее в виде ИНС-моделей можно представить как своеобразный инженерный калькулятор, позволяющий мгновенно производить требуемые вычисления характеристик горения или требуемый состав КС.

Стоит также отметить интересную роль, которую могут иметь полученные модели в учебном процессе. Обобщая закономерности экспериментальных данных, данные модели позволяют в компактном виде наглядно продемонстрировать различия КС по характеристикам горения, зависимости характеристик горения от внешних и внутренних условий, и в целом в механизме горения КС.

Все четыре полученные модели будут представлены участникам конференции для их опробования, как на основе наших данных, так и на основе данных участников конференции. При представлении участниками конференции своих данных в формате MS Excel (с числом строк - примеров наборов данных от 50 до 150 и числом столбцов – входных и выходных переменных модели от 3 до 6) могут быть получены вычислительные ИНС-модели данных участников конференции. Мы приглашаем участников конференции к участию в совместных исследованиях для решения задачи создания базы знаний в области горения КС.

---

## Литература

1. [Atwood, 1999] A. I. Atwood A. I. et al. Burning Rate of Solid Propellant Ingredients, Part 1: Pressure and Initial Temperature Effect. Journal of Propulsion and Power, Vol. 15, No. 6, 1999, p. 740-749.
2. [Makota, 2008] Makota Kohga et al. Influence of AP Particle Size on Burning Characteristics of AN/AP – Based Composite Propellants. Combustion // Advancement in Energetic Materials and Chemical Propulsion, Ed. K. K. Kuo and K. Hori, New York: Begell House, Inc., 2008, pp. 491-506.
3. [Зенин, 1980] Зенин А.А. Физические процессы при горении и взрыве. М.: Атомиздат, 1980. 69 с.
4. [Льюис, 1976] Льюис Б., Пиц Р., Тейлор Р. Процессы горения. М., Гос. издат. физической и математической литературы, 1976. с. 447 – 470.

---

## Информация об авторах



**Виктор Аbruков**, Чувашский государственный университет, Московский пр., 15 Чебоксары, 428015 Россия; e-mail: [abrukov@yandex.ru](mailto:abrukov@yandex.ru)

Области научных интересов: Data Mining, искусственные нейронные сети, горение



**Сергей Аbruков**, Чувашский государственный университет, Московский пр., 15 Чебоксары, 428015 Россия; e-mail: [abrukovs@yandex.ru](mailto:abrukovs@yandex.ru)

Области научных интересов: искусственные нейронные сети



**Елена Карлович**, Чувашский государственный университет, Московский пр., 15 Чебоксары, 428015 Россия; e-mail: [lenka-buzuluk@yandex.ru](mailto:lenka-buzuluk@yandex.ru)

Области научных интересов: искусственные нейронные сети

## PROTEIN STRUCTURE PREDICTION ON A THREE-DIMENSIONAL TRIANGULAR LATTICE

Leonid Hulianytskyi, Vitalina Rudyk

**Abstract:** *The protein tertiary structure prediction problem is examined, which is one of up-to-date problems of computational biology. The results received earlier for two-dimensional case are extended for 3D and analyzed. Algorithms of local and stochastic search, ant colony optimization are proposed. The questions of software realization and their comparative research are discussed.*

**Keywords:** *combinatorial optimization, protein structure prediction, HP model, local search, ant colony optimization.*

**ACM Classification Keywords:** *G.1.6. Mathematics of Computing, Numerical Analysis, Optimization.*

**Conference topic:** *Decision Making.*

---

### Introduction

---

Defining features of the protein molecule remains a central problem in computational biology, molecular biology, biochemistry and physics. Three-dimensional protein structure analysis is a key for understanding and manipulating its biochemical and cellular functions. Knowing the protein shape is underlying for pharmacology and medicine, as the majority of drugs' influence is caused by their conjunction to the protein for stabilizing the natural structure or destroying the part of it responsible for harmful disorders. Thus, for developing drugs it is necessary to know the exact three-dimensional structure of molecules.

Protein is a macromolecule that is composed of amino acid sequence, which number varies from 20 to 40000 (mostly about 100-300). There are 20 different amino acids. The research of proteins and their functions is usually based on "sequence-structure-function" dogma. It means that the functionality of the protein is uniquely determined by its 3D structure, while the last is defined by its amino acid sequence. High cost of experiments, amount of required time and diversity of protein forms lead to the development of mathematical methods of determining the molecule shape.

To describe the forces that affect the protein folding complex and large-scale differential equation systems are used but it is impossible to solve them in practice. Therefore, the use of simplified models that produce approximate folding is widespread. Differential equation systems are simplified on basis of these approximations, and the more exact approximation was found, the easier system becomes.

Traditionally, two different approaches are used for solving this problem. The first one, called "de novo", is based on the principles of physics, namely the thermodynamic hypothesis, according to which the natural structure of the protein corresponds to the global minimum of its free energy. Another, "template", is founded on the idea of evolution principles, assuming that similar amino acid sequences correspond to similar structures, and uses data about known folds, statistically analyzing them [Белецкий, Васильев, Гупал, 2007].

The shortcoming of "de novo"-method is the fact that the very mechanism of protein folding is not completely clear. There is a problem known as "Levintal paradox" that lies in the fact that the average protein (200 amino acids long) would be searching its secondary structure of the 3200 possible (in the secondary structure each amino acid refers to one of three forms –  $\alpha$ -helix,  $\beta$ -strand or irregular structure), each corresponding to a certain level of free energy. The protein can "feel" the stability of structure only after shaping it. The transition from one conformation to another is not faster than 10-13 seconds, so to scan all possible structures protein would spend about 1080 years (to compare, the lifetime of the Universe is 10<sup>10</sup> years) [Финкельштейн, Птицин, 2005]. Moreover, as a rule, template methods are computationally simpler and give more accurate results. However, they are not universal - if given amino acid sequence is not like any of the known, predictable precision of the evaluated fold will be small.

The simplest of "de novo"-models is the HP model offered by Dill [Dill, Bromberg, Yue, Fiebig, Yee, Thomas, Chan, 1995]. According to it 20 amino acids are divided into two groups - hydrophobic (nonpolar) and hydrophilic (polar) residues, basing on their characteristics, so the input sequence is denoted as a word from the alphabet  $\{H, P\}$ , where the symbol P indicates a polar, and H - hydrophobic residues. The idea is that while folding hydrophobic residues move inside the molecule to prevent the contact with water, and the hydrophilic remain outside. According to the hypothesis, the protein shapes a structure that minimizes the contact area of hydrophobic residues with water or polar residues. Formalizing it, every amino acid residue is located in a certain lattice node (neighboring residues in the sequence - in the neighboring nodes), contacting residues are defined as those that are not neighbors in sequence, but are located in adjacent lattice sites, and the problem is to maximize the number of contacts between hydrophobic residues. Despite the simplicity of the HP model, it properly represents real protein folding processes.

More complex models take into account the weighted influence of each contacting hydrophobic residue on the objective function [Decatur, 1996], use space discretization, based on sampling angles between amino acids, avoiding lattice binding [Awadh, Bahamish, Abdullah, Abdul Salam, 2006], base on protein secondary structure, operating with relative positions of  $\alpha$ -helices,  $\beta$ -strands and irregular structures [Paluszewski, Winter, 2008].

Various methods of combinatorial optimization are applied for solving the protein tertiary structure prediction problem in the HP model, such as simulation annealing algorithm, evolutionary algorithms, ant colonies optimization algorithms and others. Though the problem remains acute: open questions are improving the model (e.g. choosing more suitable lattice) and optimizing the time and accuracy of computing.

---

### HP-model

---

In the current research for prediction of the protein tertiary structure Dill's HP-model was chosen, as the most common one. Let's describe it in detail.

Each of the 20 amino acids applies to one of the two classes - hydrophobic ( $H$ ) or polar ( $P$ ), according to their physical properties, so amino acid sequence can be denoted as a sequence  $S = \xi_1 \xi_2 \dots \xi_n$ ,  $\xi_i \in \{H, P\}$ ,  $i = \overline{1, n}$  of length  $n$ . Each residue is placed in a node of some discrete lattice (such correspondence is called fold) in such a way that adjacent elements in a sequence correspond to neighboring nodes of the lattice (connectivity property), defining self-avoiding path on it. Connected self-avoiding folding is considered to be feasible. According to Dill contact appears between those hydrophobic residues that are not adjacent in the sequence, but are located in neighboring nodes of the lattice. The energy of folding is calculated as a number of contacts in it with a negative sign

More formally, if each element of the amino acid sequence assigned to some node, the energy is calculated as follows:

$$E(S) = - \sum_{1 \leq i < j \leq n-2} I(L_i, L_j) h(\xi_i, \xi_j) \quad (1)$$

where

$$I(L_i, L_j) = \begin{cases} 1, & \text{if nodes } L_i \text{ and } L_j \text{ are neighbors,} \\ 0 & \text{else,} \end{cases}$$

$$h(\xi_i, \xi_j) = \begin{cases} 1, & \text{if both } \xi_i \text{ and } \xi_j \text{ are hydrophobic,} \\ 0 & \text{else.} \end{cases}$$

The protein tertiary structure prediction is reduced to finding conformation with the lowest energy. In such interpretation the problem is proved to be NP-hard [Berger, Leighton, 1998; Crescenzi, Goldman, Papadimitriou, Piccolboni, Yannakakis, 1998].

Initially for 2D case Dill used square lattice. But one of its significant drawbacks is that the two residues can be adjacent in the lattice only when the number of elements in the sequence between them is even (it is known as parity problem). Thus, the sequence  $(HP)^n$  on a square lattice will have no connections. That fact contradicts the natural representation. Moreover, such a strict limitation complicates the analysis of approximate algorithms, identifying artificially high lower energy bound, resulting in the fact that the algorithm accuracy estimation won't have any significant value in real problems. Therefore in current research the triangular lattice was chosen for two-dimensional case and its axonometric variant in case of three dimensions (Fig. 1) where there is no parity problem, as for any two positions in the sequence there exists a conformation with the corresponding residues assigned to neighboring nodes of the lattice.

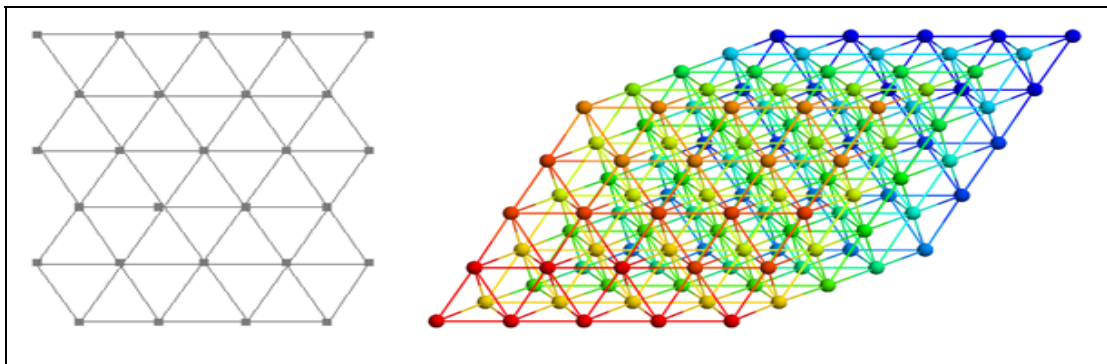


Fig. 1. Two-dimensional and axonometric triangular lattices.

### Problem definition on three-dimensional lattice

Let  $\xi_1 \xi_2 \dots \xi_n$ ,  $\xi_i \in \{H, P\}$  indicate the sequence of amino acids of length  $n$ . To formulate the problem as the COP first of all it is necessary to define the decision space, in other words a fold should be denoted as some mathematical object [Гуляницкий, Рудык, 2010].

The simplest way to do it is to encode a folding as a sequence  $c_1 c_2 \dots c_n$  of integer coordinates of each amino acid residue so that a point  $c_i = (x_i, y_i, z_i)$ ,  $x_i, y_i, z_i \in \mathbb{Z}$  assigns to a lattice node with Cartesian coordinates  $x_i \mathbf{e}_1 + y_i \mathbf{e}_2 + z_i \mathbf{e}_3$ , where  $\mathbf{e}_1 = (\frac{\sqrt{3}}{2}, \frac{1}{2}, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$ ,  $\mathbf{e}_3 = (\frac{\sqrt{3}}{6}, \frac{1}{2}, \frac{\sqrt{33}}{6})$ .

In such notation nodes  $(x_1, y_1, z_1)$  and  $(x_2, y_2, z_2)$  are neighbors if and only if  $(x_1 - x_2, y_1 - y_2, z_1 - z_2) \in A$ ,

$$A = \{ (0, 1, 0), (1, 0, 0), (1, -1, 0), (-1, 0, 0), (-1, 1, 0), \\ (0, -1, 0), (0, 0, 1), (0, -1, 1), (-1, 0, 1), (0, 1, -1), (1, 0, -1), (0, 0, -1) \}.$$

This representation is called coordinate encoding and it is comfortable for checking if a fold is a self-crossing one, but in general case changing a single sequence element leads to losing a connectivity property. That complicates the process of algorithm development. Another drawback is that the same fold corresponds to some different assignments turning one into another by a turn or parallel transport. To exclude the equal variants from examination the additional constraints that are difficult to hold are imposed.

For avoiding these drawbacks another encoding, called absolute, is proposed. A fold is denoted as a sequence  $a_1 a_2 \dots a_{n-1}$ ,  $a_i \in A$ ,  $a_i = c_{i+1} - c_i$ , it fixes the position of each amino acid residue (starting from second) with respect to the previous one. Every such code specifies a connected path in the lattice, and for checking self-avoidance constraint it is necessary to convert a sequence to coordinate encoding. Partially the problem with ambiguity is solved - parallel transport of fold does not change its code, but turns in various directions lead to existence of 12 different codes of the same conformation.

The other way to present a fold is to describe it by relative encoding. It represents a position of every residue depending on a part of foregoing fold, in such a way it fixes angles in the conformation. So in case of two-



dimensional lattice six elements of absolute encoding represent six possible directions: "west", "northwest", "northeast", "east", "southeast" and "southwest", while there are only five in relative encoding and they mean "back-left", "front-left", "front", "forward-right", "back-right" [Гуляницкий, Рудык, 2009] (Fig. 2).

This representation has following advantages:

- each convolution corresponds to its relative code one-to-one;
- the absence of "back" partially direction solves a problem with self-crossing in a fold.

A feature is the fact that the modification of single sequence element of relative code leads to a turn of following part of the fold, while in absolute encoding this action parallel transfers the same part. Thus comparing with absolute encoding a small change of code causes more significant change of fold energy value, but there is higher probability for self-crossing to appear.

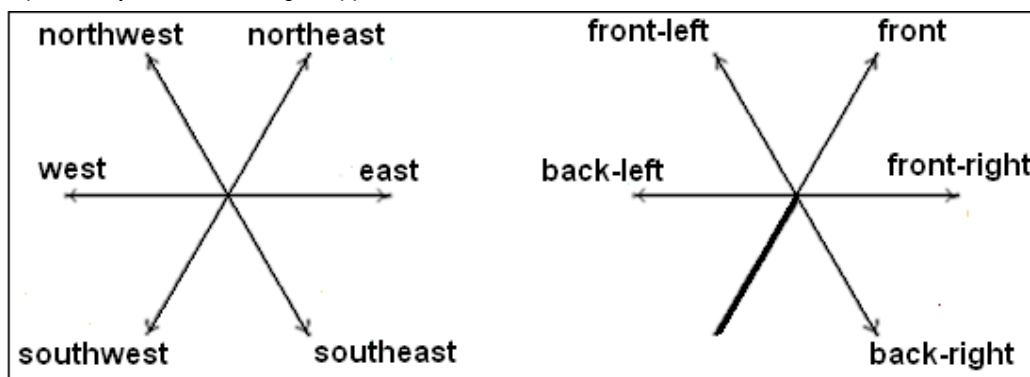


Fig. 2. The elements of absolute and relative code in two-dimensional triangular lattice.

The disadvantage is the complexity of conversion a relative code to coordinate to check self-avoidance: in two-dimensional lattice relative code element depends on the current and previous absolute code elements. (So if the last absolute direction was "east", the next relative direction "front-right" will match the absolute "southeast"). In axonometric case it depends on previous two. To convert the absolute code to relative and vice versa a conversion table is built. Its first two columns are filled with possible absolute directions, the third applies to a desired relative one, and the fourth gives a resulting absolute.

Using the relative encoding a fold is formally represented in the following way:

$$r_1 \in \tilde{R}, r_i \in R, i = \overline{2, n-2}, R = A \setminus \{(0, -1, 0)\},$$

$$\tilde{R} = \{(0, 1, 0), (1, 0, 0), (1, -1, 0), (1, 0, -1)\}.$$

After comparing three encodings two of them (absolute and relative) were chosen for further analysis. Thus COP is defined: the encoding determines a decision space, a set of feasible solutions is the set codes that represent a self-avoiding fold, and the objective function is the energy of fold determined by formula (1).

### Deterministic local search

Local search algorithms are a group of algorithms which search optimal values of objective function by iteratively moving from one acceptable solution to another neighboring one, and stop when certain conditions are met (e.g. time or iteration number constraints). The input of the algorithm is some initial value which serves as search starting point. Usually these algorithms are fast and significantly improve initial solution, but in case of multicriterion objective functions they rarely find global optima. These features allow them to be used as a part of more complex algorithms.

Let's examine the deterministic local search algorithm for protein energy minimization in HP model. The distance  $\rho(s, v)$ , between folds  $s = (s_1, s_2, \dots, s_k)$ ,  $v = (v_1, v_2, \dots, v_k) \in D$  is determined as the number of differing characters in their encodings:

$$\rho(s, v) = \sum_{i=1}^k \chi(s_i, v_i), \quad \chi(x, y) = \begin{cases} 1, & x = y; \\ 0, & x \neq y. \end{cases}$$

Let  $O_\delta(s)$  denote a fold  $s$  neighborhood of size  $\delta$ , so  $O_\delta(s) = \{v \mid \rho(s, v) \leq \delta\}$ . Using such notation the deterministic local search scheme is given in Fig. 3.

```

procedure LocalSearch (  $s_0$  )
  while  $O_1(s_0)$  is not totally examined do
     $s :=$  some fold from  $O_1(s_0)$ ;
    if  $s \in D$  and  $E(s) < E(s_0)$  then  $s_0 := s$ ;
  end while;
  return  $s_0$ ;
end procedure.

```

Fig. 3. Deterministic local search algorithm scheme.

Thus the algorithm guarantees the improvement of current fold in neighborhood in case it exists.

With the help of deterministic local search the difference between absolute and relative encoding can be demonstrated. As it was mentioned above, a change of single character in absolute encoding leads to parallel shift of the succeeding molecule part, while in relative it results in rotation. Fig. 4 shows the folds produced by the algorithm which was applied to a fold that had a line shape (using relative encoding – on top, and using absolute – on bottom).

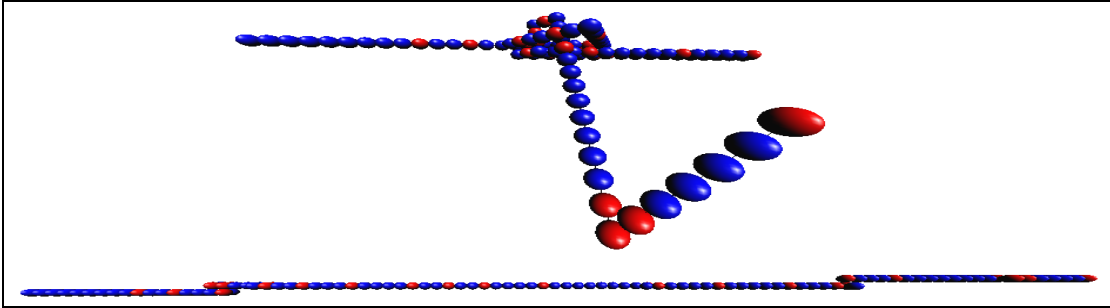


Fig. 4. Deterministic local search results.

The problem is that the condition of self-avoidance is rather strict, so the neighborhood  $O_1(s)$  can contain few to none acceptable solutions. Taking this into account it is proposed to replace neighborhood  $O_1(s)$  in the procedure LocalSearch by a neighborhood of bigger size. It nonlinearly increases the time of every iteration: the maximum number of elements in  $O_1(s)$  is  $k * d$ , where  $k$  denotes the length of encoding and  $d$  means the number of characters in it ( $d=12$  for absolute encoding and  $d=11$  for relative one) while  $O_2(s)$  may contain  $k(k-1) * d^2$  folds. Therefore it is reasonable to reduce  $O_2(s)$  to  $\tilde{O}_2(s)$  according to the rule

$$v \in \tilde{O}_2(s) \Leftrightarrow \exists p \in N, 1 \leq p \leq k-1: v_t = s_t \forall t \in \{1, \dots, k\} \setminus \{p, p+1\},$$

and to use it instead of  $O_1(s)$  (maximum number of elements in  $\tilde{O}_2(s)$  is  $(k-1) * d^2$ ), although it also contributes greatly to the time. Therefore, to improve performance it is proposed to use stochastic local search methods.

#### Simulated Annealing Algorithm

Simulated annealing algorithm is one of the widely used local optimization algorithms [Hoos H, Stützle, 2005]. Its idea is quite similar to the one that deterministic local search has, but due to the stochastic component the

algorithm may recover from local minima, returning more accurate value. The basic idea is that the move to the best value is executed always, while moves to the "worse" (in terms of objective function values) are accepted with a certain probability that depends on the objective function value and iteration number.

The algorithm is based on simulation of physical process taking place during the crystallization of substances from liquid to solid state, particularly during annealing. It is believed that atoms have taken their stable places in crystal lattice, but transitions of individual atoms from one node to another are still possible. The process proceeds under conditions of gradually lowering temperature. The atom transition from one place to another occurs with some probability that is decreasing with decrease of temperature. Stable crystal lattice corresponds to the minimum of atoms energy, so the atom tends to move to a state with lower energy.

The scheme of the algorithm is shown in Fig. 5.

The temperature varies according to the rule

$$T_t = \frac{H}{\log\left(\frac{t}{P}\right)},$$

where  $H, P \in R$  are the algorithm parameters.

The equilibrium condition is defined as follows. Let  $\nu$  be some natural number and  $\varepsilon > 0$  is real.  $\nu$  successive iterations of the algorithm are called "run". If on the current temperature level  $q$  runs were made, it is presumed that the equilibrium condition is met if

$$|E_{q+1} - E_i| \leq \varepsilon$$

for some  $i \in \{1, \dots, q\}$ , where  $E_i$  is the average energy value on  $i$ -th run.

```

procedure SA(  $s_0$  )
   $s_{rec} := s_0$ ;  $T :=$  some initial value;
  while  $O_1(s_0)$  is not totally examined do
    while equilibrium condition is not met do
       $s :=$  some fold from  $O_1(s_0)$ ;
      if  $s \in D$  then
         $\Delta := E(s) - E(s_0)$ ;
         $p := \min\{1, \frac{\Delta}{T}\}$ ;
         $\xi :=$  random[0,1];
        if  $p \geq \xi$  then
           $s_0 := s$ ;
          if  $E(s) < E(s_{rec})$  then  $s_{rec} := s$ ;
        end if;
      end if;
    end while;
     $T :=$  next value;
  end while;
  return  $s_{rec}$ ;
end procedure.

```

Fig. 5. Simulation annealing algorithm scheme

### G- algorithm

G-algorithm is an efficient algorithm for solving combinatorial optimization problems [Гуляницкий, 2004]. Unlike the simulation annealing algorithm, the probability of transition to the "worse" value on each step does not depend on the number of iteration, but the bound value that defines dropout conditions for "worse" folds does (Fig. 6).

The function  $\varphi(x, y)$  is defined as follows:

$$\varphi(x, y) = 1 - \frac{E(y) - E(x)}{\gamma \cdot (E(x) - E_{\min})},$$

where  $E_{\min}$  denotes the lower bound of protein energy and  $\gamma > 0$  is a parameter. For evaluating  $E_{\min}$  the following property is used: each hydrophobic amino acid residue except the first and the last ones in a triangular lattice can form up to 10 contacts, while the first and the last ones can have 11. Therefore, if  $n_H$  denotes the number of hydrophobic residues in amino acid sequence,  $E_{\min}$  can be calculated by the formula:

$$E_{\min} = -\frac{10n_H + 2}{2} = -(5n_H + 1).$$

**procedure G-algorithm ( $s_0$ )**

$s_{rec} := s_0$  ;  $\mu_0 := 0$  ;  $h := 0$  ;  $t := 0$  ;

**while**  $O_1(s_0)$  **is not totally examined do**

**while** equilibrium condition **is not met do**

$s :=$  some fold from  $O_1(s_0)$  ;

**if**  $s \in D$  **then**

$\tilde{\varphi} := \varphi(s_0, s)$  ;

$p := \min\{1, \max\{0, \tilde{\varphi}\}\}$  ;

$\xi := \mu_t + \text{random}[0, 1] \cdot (1 - \mu_t)$  ;

**if**  $p \geq \xi$  **then**

$s_0 := s$  ;  $h := h + 1$  ;

**if**  $E(s) < E(s_{rec})$  **then**  $s_{rec} := s$  ;

**end if** ;

**end if** ;

**end while** ;

$\mu_{t+1} := G(\mu_t)$  ;  $t := t + 1$  ;

**end while** ;

**return**  $s_{rec}$  ;

**end procedure.**

Fig. 6. G-algorithm scheme.

Function  $G(x)$  was chosen as follows:

$$G(x) = (\eta_h(x) + H)^h,$$

where  $0 < H < 1$ ,  $h \in \{1, 2, 3\}$  are the algorithm parameters,

$$\eta_k(x) = \begin{cases} 0, & x = 0, \\ x^{1/h}, & x > 0. \end{cases}$$

Equilibrium condition is the same as one used for the simulated annealing algorithm.

Simulated annealing algorithm and G-algorithm in general give more significant improvement of the initial solution than deterministic local search, and the calculation time increases slightly, what makes them suitable for further use.

#### Ant Colony Optimization Algorithm

Ant colonies optimization algorithm belongs to up-to-date combinatorial optimization stochastic methods and is widely spread [Dorigo M., Stützle, 2004]. The essence of the approach is based on analyzing and using ants' behavior model when they seek shorter paths from colony to food.

The main idea of the algorithm takes as a principle formic colony behavior, such as marking the shorter routes with bigger number of pheromones. In every iteration fixed number of ants simultaneously search their way to the food, using probability based methods of choosing directions that depend on the length of the relevant section of the route and the number of pheromones in it. While traveling some path an ant marks it with the number of pheromones, proportional to its length. To gradually remove nonoptimal plots pheromones evaporation procedure is used. Currently, these methods are quite competitive compared to other metaheuristics and to date they even produce the best results for some problems.

The general algorithm scheme is shown in Fig. 7, where  $h, p \in \mathbb{N}$  are the algorithm parameters.

The dimension of pheromone matrix for the examined problem is  $d \times (n-2)$  (as was mentioned before,  $d$  means the number of characters in the encoding,  $d = 12$  for absolute and  $d = 11$  for relative one). In the first step the matrix can be initialized randomly or with certain given value  $m_0 \in \mathbb{R}^+$ . At each iteration the probability follow  $i$ -th direction after  $(j+1)$ -th residue in the structure is proportional to  $[i, j]$ -th matrix element to the power  $\alpha$ ,  $\alpha \in [0, 1]$  and position estimation  $e(pos)$  to the power  $1-\alpha$ , that is calculated by the following rule:

$$e(pos) = \sum_{p \in O(pos)} e_{\xi_{j+2} f(p)},$$

```

procedure ACO()
   $s_{rec} := null$ 
  InitPheromoneMatrix (M);
  while stopping condition is not met do
    for  $i = 1, \dots, h$  do
       $s_i := \text{GetSomeFeasibleSolution}(M)$ ;
      if  $i \leq p$  then  $s_i := \text{LocalSearch}(s_i)$ ;
      if  $E(s_i) < E(s_{rec})$  then  $s_{rec} := s_i$ ;
    end for;
    for  $i = 1, \dots, h$  do RefreshPheromoneMatrix(M,  $s_i$ );
  end while;
  return  $s_{rec}$ ;
end procedure.

```

Fig. 7. Ant colony optimization algorithm scheme.

where  $pos$  is a lattice node, which the next residue may be positioned in,  $O(pos)$  is a set of its neighboring nodes,  $f(\cdot) \in \{H, P, 0\}$  is a function that returns  $H$  if the node is occupied by hydrophobic residue,  $P$  in case of polar and  $0$  if the node is free. Numbers  $e_{HH}, e_{HP}, e_{H0}, e_{PH}, e_{PP}, e_{P0} \geq 0$  are the algorithm parameters and should satisfy following conditions:  $e_{HP} \leq e_{H0} \leq e_{HH}$ ,  $e_{PH} \leq e_{P0} \leq e_{PP}$ .

It is proposed to refresh the pheromone matrix using one of two ways. The first of them is to increase each element proportionally to the absolute value of generated fold according to the formula

$$m_{ij}(t+1) = (1-\rho)m_{ij}(t) + \sum_{l=1}^h \chi_{i,j+1}(s_l)(-E(s_l))^q,$$

where  $m_{ij}(t)$  is  $[i, j]$ -th matrix element on the step  $t$ ,  $q \in R^+$  is the algorithm parameter,  $\rho \in [0, 1)$  defines the rate of pheromones evaporation,  $\chi_{i,j}(s) \in \{0, 1\}$  is a characteristic function that equals 1 if and only if the encoding of fold  $s$  has its  $j$ -standing position occupied by symbol  $i$ .

The second way is notable for considering separately those minimum parts of molecules, which have contacts inside:

$$m_{ij}(t+1) = (1-\rho)m_{ij}(t) + \sum_{l=1}^h \chi_{i,j+1}(s_l)(Z_{j+1}(s_l))^q,$$

where  $Z_j(s)$  is a number of contacts in fold  $s$  between hydrophobic amino acid residues that have an order not greater than  $j$ , and ones with order greater than  $j$

A local search procedure can be selected from three algorithms described above.

#### Computational Experiment

The efficiency of introduced algorithms was analyzed with the help of specially designed software tools that consist of 3 programs:

- algorithm visualizer is a program that allows to analyze step-by-step the chosen algorithm with the help of charts that demonstrate dependency of energy from iteration number, to customize its input parameters, to observe the energy of output protein folding and it's 3d shape;
- data generator is a software designed for automatic maintenance of database that stores the output of protein structure prediction algorithms;
- data analyzing module is program designed for visual analysis of algorithms and their comparison one with another or the same ones with different values of parameters.

The computational experiment was conducted using the database of real proteins SwissProt [SwissProt]. For each local search algorithm 20 problems were solved using absolute and relative encoding and fixed input parameter values. Each of them started from three different randomly generated folds and a fold that has a line shape. For each stochastic algorithm for a fixed set of "amino acid sequence-initial folding" 5 runs were conducted. Thus, for each deterministic local search algorithm 80 launches were made, and for simulated annealing and G- algorithm – 400 launches. Using ACO algorithms with fixed set of parameters 5 problems were solved, 3 times each. To compare the effectiveness of the algorithm with different parameter values the additional experiments were carried out.

Folds that were used as the initial approximations for algorithms and the starting seeds of pseudorandom numbers generator are saved in the database to make computational experiment reproducible.

The outcomes of the computational experiment are the following:

Local search algorithms that use absolute encoding in general improve the initial folding better than those, which use relative, except the case it has a shape of line. It can be explained by the fact that varying the single element of relative encoding rotates succeeding part of fold, so it noticeably changes, but at the same time some contacts are saved. The fold that has a line shape is flexible for such actions, but compact folds are not, as their neighborhood in relative encoding in general contains much less feasible solutions.

The increasing of neighborhood radius in deterministic local search algorithms leads to a significant improvement of initial solution, but it nonlinearly increases expended time too. The same is correct for comparing stochastic local search methods with determined local search using single radius and comparing determined local search using extended radius with stochastic local search algorithms. Methods that use a local search procedure for the

improvement of somehow generated folds (e.g. genetic algorithms, ACO) should balance using different local search methods to achieve more accurate results during some reasonable time.

On the average solutions generated by G-algorithm are more accurate than those, which were received by simulated annealing with the same time expended.

The efficiency of simulated annealing, G-algorithm and ant colony optimization algorithm considerably relies on parameter values, what confirms the need to develop automatic methods of choosing appropriate parameter values.

The described above features of relative encoding develop into evident advantage in both accuracy and time compared to absolute encoding when using them in ant colony optimization algorithm.

Using position estimation and partial refreshing of pheromone matrix in ant colony optimization algorithms both improve its quality reasonably increasing expended time, whereas the improving of generated folds on every iteration with the help of local search procedure does not result in perfection of derived solution as it was expected. It can be explained by two facts. First of all, using local search for some folds on every step leads to enormous expenditure of time, what considerably decreases the algorithm speed. And the second reason is that significant improvement of generated folds results in premature [convergence](#). From the other side, without local search ant colony optimization algorithm remains uncompetitive, so the open question is varying the parameters and adjusting the algorithm for the use of local search to be effective.

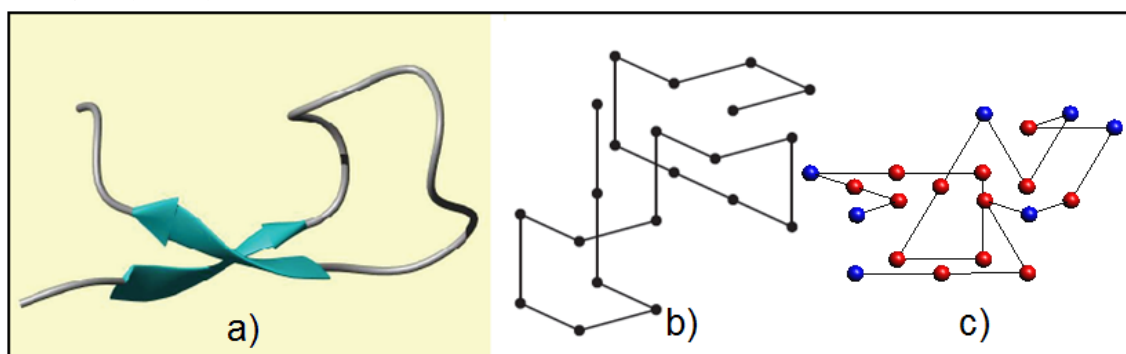


Fig. 8. Hepsidin.

To compare received results with the real ones, the problem of predicting the tertiary structure of Hepsidin molecule (the corresponding amino acid string is HPPHPPPPHPPHPPHPPH) was solved. The results are shown in Fig. 8. The image a) corresponds to Hepsidin real known structure, b) was predicted by ACO method in cubic lattice [Fidanova, Lirkov, 2008], and c) was generated by the proposed ACO method with relative encoding in three-dimensional triangular lattice. The visual comparison on the one hand confirms the adequacy of the HP model and the feasibility of selecting a triangular lattice, on the other it shows the necessity of new sophisticated models' development which more accurately take into account the traits of folding process, and the development of comparison methods to estimate the similarity between the discrete fold and the real one.

## Conclusion

The combination of different fold encodings with well-known combinatorial optimization methods results in creating competitive algorithms for solving protein tertiary structure prediction problem. The detailed analysis forms a base for their former use as a part of complex metaheuristic algorithms.

Open questions are the theoretical research of ant colony optimization algorithms convergence and combining them with local search methods, the development of hybrid algorithms for solving the problem, researching new models that describe folding processes in more precise way.

---

## Благодарности

*The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria [www.ithea.org](http://www.ithea.org), and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine [www.aduis.com.ua](http://www.aduis.com.ua).*

---

## Бibliography

- [Белецкий, Васильев, Гупал, 2007] Белецкий Б.А., Васильев С.В., Гупал А.М. Предсказание вторичной структуры белков на основе байесовских процедур распознавания // Проблемы управления и информатики. – 2007. – №2. – С.59-64.
- [Гуляницкий, 2004] Гуляницкий Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – 2004. – №1. – С. 64-72.
- [Гуляницкий, Рудык, 2009] Гуляницкий Л.Ф., Рудык В.А. Разработка и исследование алгоритмов решения задачи прогнозирования третичной структуры протеина // XVth International Conference «Knowledge-Dialogue-Solution» KDS 2009 – С. 104-112.
- [Гуляницкий, Рудык, 2010] Гуляницкий Л.Ф., Рудык В.А. Моделирование свертывания протеина в пространстве // Компьютерная математика. – 2010. – №1. – С. 128-138.
- [Финкельштейн, Птицин, 2005] Финкельштейн А.В., Птицин О.Б. Физика белка: Курс лекций с цветными и стереоскопическими иллюстрациями и задачами. – 3-е изд., испр. и доп. – М.: КДУ, 2005. – 456 с.
- [Awadh, Bahamish, Abdullah, Abdul Salam, 2006] Awadh H. Bahamish A., Abdullah R., Abdul Salam R. Protein conformational search using honey bee colony optimization // Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications. – 2006.
- [Berger, Leighton, 1998] Berger B., Leighton T. Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete // Journal of Computational Biology. – 1998. – № 5(1). – P. 27-40.
- [Crescenzi, Goldman, Papadimitriou, Piccolboni, Yannakakis, 1998] Crescenzi P., Goldman D., Papadimitriou C., Piccolboni A., Yannakakis M. On the complexity of protein folding // Journal of Computational Biology. – 1998. – № 5(3). – P. 423-465.
- [Decatur, 1996] Decatur S.E. Protein folding in the generalized hydrophobic-polar model on the triangular lattice. MIT LCS Technical Memo: MIT-LCS-TM-559. – 1996.
- [Dill, Bromberg, Yue, Fiebig, Yee, Thomas, Chan, 1995] Dill K., Bromberg S., Yue K., Fiebig K.M., Yee D., Thomas P., Chan H. Principles of protein folding – a perspective from simple exact models // Protein Science. – 1995. – № 4. – P. 561-602.
- [Dorigo M., Stützle, 2004] Dorigo M., Stützle T. Ant Colony Optimization. – Cambridge: MIT Press, MA, 2004. – 348 p.
- [Fidanova, Lirkov, 2008] Fidanova S., Lirkov I. Ant colony system approach for protein folding // Proceedings of the International Multiconference on Computer Science and Information Technology. – 2008. – P. 887-891.
- [Hoos H, Stützle, 2005] Hoos H., Stützle T. Stochastic Local Search: Foundations and Applications. – San Francisco: Morgan Kaufmann Publ., 2005. – 658 p.
- [Paluszewski, Winter, 2008] Paluszewski M., Winter P. EBBA: Efficient Branch and Bound Algorithm for Protein Decoy Generation // DIKU, Technical Report. – 2008.
- [SwissProt] SWISS\_PROT Protein Sequence Data Bank: <http://www.expasy.ch/sprot/>

---

## Authors' Information



**Leonid Huliannytskyi** – Dr.Sc.(Technology), Leading Research Scientist of V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine; Prof. of NTUU "KPI"; e-mail: [lh\\_dar@hotmail.com](mailto:lh_dar@hotmail.com)

*Major Fields of Scientific Research: Combinatorial optimization, Decisions Making, Mathematical modeling and practical applications*



**Vitalina Rudyk** – Master of Science in Applied Mathematics at Taras Shevchenko National University of Kyiv, Ukraine; e-mail: [vitalina.rudyk@gmail.com](mailto:vitalina.rudyk@gmail.com)

*Major Fields of Scientific Research: Combinatorial optimization and practical applications, Computational biology, Protein folding*



---

---

## Decision Making

---

---

### INDIVIDUALLY OPTIMAL PRINCIPLES OF DISTRIBUTION OF GREENHOUSE GAS EMMISSION QUOTAS

Olexij Voloshin, Sergij Maschenko

**Abstract:** *The problems of climate change mitigation are being analyzed. Different emission quotas distribution mechanisms are observed. The work offers model and method of distribution based on the individual rationality principle.*

**Keywords:** *climate protection, emissions of greenhouse gases, global limit on emissions, distribution methods, Nash equilibrium, individual optimal equilibrium.*

**ACM Classification Keywords:** *I. Computing Methodologies – I.6. Simulation and modelling (Time series analysis) – I.6.8 Types of Simulation – Gaming.*

---

#### Introduction

This work continues the research presented in [Voloshin, 2009,2010]. Climate changes caused by anthropogenic influence on the Earth's atmosphere are one of the most important environmental problems of civilization. The first step in solving this problem was the signing by 154 countries in 1992, the UN Framework Convention on Climate Change (UNFCCC). In 1997 "The Kyoto Protocol to the United Nations Framework Convention on Climate Change" was adopted. Under the Protocol, 37 industrialized countries (called "Annex I countries") commit themselves to a reduction of four greenhouse gases (GHG) (carbon dioxide, methane, nitrous oxide, sulphur hexafluoride) and two groups of gases (hydrofluorocarbons and perfluorocarbons) produced by them, and all member countries give general commitments. Also it provides mechanisms for "trading" emissions. The problem lies in the fact that less than half the countries signatories of the UNFCCC, ratified the Kyoto Protocol (KP). In particular, the KP wasn't signed by the industrial giants - the U.S., China and India. They argue, that KP offers "unfair" principles for assigning emission rights. Discussion of these issues over the last decade has not led to constructive solutions. In particular, the results of the last UN conference on climate changes in December 2009 in Copenhagen proved to be disappointing. The Berlin conference in May 2010, involving representatives from over 100 countries, also did not lead to consensus. In [Voloshin, 2009,2010] statistics on the problem are provided, criteria and indicators, which are accounted (can be taken into account) in the allocation of quotas and the causes of "dead end" situation in solving these problems are analyzed. In [Voloshin, 2009] task of quotas allocation is reduced to the classical problem of cost allocation [Voloshin, 2006], which, by turn, is considered as a cooperative game. The analysis in [Voloshin, 2010] on the basis of the results of recent talks in Copenhagen and Berlin, suggests that the desire of participants to cooperation (regional level, level of development, etc.) is missing. In particular, China and India are ready to participate in activities to implement the KP, if their share of the quotas will be determined based on the amount of emissions per capita. In [Voloshin, 2010] is proposed: first, to base the quotas allocation on the concept of multi-objective optimization using the maximum number of available indicators of economical, environmental, demographical and social development of countries and, secondly, to use a two-tiered procedure for finding a compromise: at the first level the allocation mechanisms have to be agreed, at the second level received allocations have to be analyzed for addressing their integral "quality" which is expressed by the utility function.

### Game Models of Distribution

The last meeting in Copenhagen showed the problem of KP can be the most adequately described by a model of decision making in conditions of complete information with minimal co-operation of players (ideally - in the total absence of cooperation) [Voloshin, 2006]. In such models, the most reasonable is the concept of "Nash equilibrium" - a situation when no player can benefit by changing his or her strategy while the other players keep theirs unchanged. And the most interesting thing is that (as we believe) the world community is in a similar situation! The foregoing is illustrated by a simple example of Samuelson [Samuelson, 2008]. For convenience we will change the numerical values of "profits" of players (without loss of content). Let us consider the game, described in Table 1, where L<sub>1</sub>, L<sub>2</sub> – respectively, the strategy of players 1 and 2, targeted at the low level of contamination; H<sub>1</sub>, H<sub>2</sub> – high level. The cells (situations) A-D show the profit of 1 and 2 players. The only non-cooperative Nash equilibrium (cell D) is inefficient – none of the agents are not able to increase profits by reducing the level of pollution. A shift in cell A (with the same profit players, but with lower levels of contamination) is possible (by Samuelson) by the intervention of "government" (coordinating body) in the establishment of the fines or quotas on emissions. But no government can help, if we change the profit values in cell D, for example, at (100, 100)!

Table 1

1 \ 2	L <sub>2</sub>	H <sub>2</sub>
L <sub>1</sub>	A <b>10; 10</b>	B <b>-5; 15</b>
H <sub>1</sub>	C <b>15; -5</b>	D <b>10; 10</b>

Table 2

1 \ 2	L <sub>2</sub>	H <sub>2</sub>
L <sub>1</sub>	A <b>20; 20</b>	B <b>-5; 15</b>
H <sub>1</sub>	C <b>15; -5</b>	D <b>10; 10</b>

Table 3

1 \ 2	L <sub>2</sub>	H <sub>2</sub>
L <sub>1</sub>	A <b>10; 7</b>	B <b>0; 8</b>
H <sub>1</sub>	C <b>8; 0</b>	D <b>1; 1</b>

Which way out of this situation? First, to obtain "objective" values of "wins" (calculated by taking into account the parameters given in [Voloshin, 2010]), and secondly, it is necessary to depart from the concept of "non-cooperative Nash equilibrium". Rationale for the last assertion comes from the following simple generalization of Samuelson model (Table 2). In this game there are two Nash equilibrium (cells A and D), and in equilibrium A win for each player 2 times greater (with low pollution) and it would seem that the agents could agree on the choice of equilibrium A. However, if in situation A the risk of rejection of each of the players is very high (only losing with 25% gains, while bringing losses to a competitor), then deviation from the situation of D, in principle, is not beneficial to every player. The concept of Nash equilibrium, taking in count "the efficiency of gain" and "the effectiveness of risk" is described in [Harsanyi, 2001]. But this approach does not significantly change the situation of decision making in the KP implementation. It is shown by the following example (Table 3). There is only one Nash equilibrium (cell D), but it is "very ineffective". Cell A gives each player gains substantially greater, but for the player 2 is profitable to deviate from the situation A. How to "force" players choose the situation A? The answer – each player must take into account the interests of another. Formally, the concept of "compromise for the sake of resolving the conflict" (each player chooses a strategy individually, taking into account the interests of other agents) developed in [Maschenko, 2007, 2009], where the principle of "individual optimality" is proposed. It is proved primarily for single-purpose games in which all agents have one goal, but for each agent it is characterized by its payoff (utility) function.

### Individually optimal allocations

The problem of collective decision-making in the following statement is being observed:

$$\begin{aligned} u_i(x) &\rightarrow \max, i \in N = \{\overline{1, n}\}, n \geq 2, \\ x \in X &= \prod_{i=1}^n X_i \end{aligned} \quad (1)$$

where  $u_i$  – a scalar utility function of  $i$ -th agent;  $X_i$  – set of its strategies;  $x = (x_1, \dots, x_n)$  – a situation the decision-making.

The principle of individual optimality is based on the following relation of domination in Nash sense.

*Definition 1.* Situation  $y = (y_i, x_{N \setminus i})$  is in relation of strong domination in Nash sense for agent  $i \in N$ , if

$$u_j(y_i, x_{N \setminus i}) > u_j(x), \forall j \in N. \quad (2)$$

Note that in (2) the effect of changes in strategy  $i$ -th player (from  $x_i$  to  $y_i$ ) to the values of utility functions of all players (in particular,  $i$ -th, which determines the Nash equilibrium) is analyzed.

*Definition 2.* The situation  $x^*$  is called weak individually-optimal equilibrium, if there is no such player  $i \in N$  and situation  $x \in X$  that would strongly dominate  $x^*$  in Nash sense.

Application of weak individually-optimal equilibrium is motivated by next scenario. Players enter into an optional agreement (the players listen to scientific advice of experts) that they would adhere to the situation  $x^*$ . If and only if the base of agreement is a weak individually optimal strategy, changes of  $i$ -th agent ( $i \in N$ ) strategy  $x_i^*$  that is agreed with the other players to any other, will always lead to a situation which would not be better for at least one player. As this situation cannot be the best for all players at the same time, then it is a compromise.

In [Maschenko, 2009] necessary and sufficient conditions for individual optimality, criterion of "stability" of the situation, the principles of choice of individually-optimal equilibrium are observed.

We assume that the payoff function of all the players are limited to a variety of situations  $X$  and denote  $S = \sum_{j \in N} \sup_{x \in X} u_j(x) < \infty$  - the sum of their upper limits. Consider for each player  $i \in N$  a set of parameters

$$M_i = \{\mu_i = (\mu_i^j)_{j \in N} \mid \sum_{j \in N} \mu_i^j = 1; \mu_i^j \geq 0, j \in N\}.$$

Necessary and sufficient conditions for individual optimality are established in the next theorem.

**Theorem 1.** If the situation  $x^*$  is individually-optimal, utility functions  $u_j(x^*) > 0, j \in N$ , then there are vectors of parameters  $\mu_i \in M_i, i \in N$ , in particular with the components:

$$\bar{\mu}_i^j = u_j(x^*) / S - \sum_{k \in N} u_k(x^*) / S n, j \in N, \quad (3)$$

that for  $\forall x_j \in X_j, i \in N$  the next inequalities are true:

$$\min_{j \in N} (u_j(x_i, x_{N \setminus i}^*) - S \mu_i^j) \leq \min_{j \in N} (u_j(x^*) - S \mu_i^j), \quad (4)$$

Any solution  $x^*$  of inequalities (4) for given  $\mu_i \in M_i, i \in N$  is weak individually-optimal equilibrium.

It should be noted that the parameters  $\mu_i \in M_i$  allow the player to express their preference on the set of payoff functions of all players  $u_j(x), j \in N$ . For example, if he considers that the function with index  $j_1 \in N$  is more important for him than the one with index  $j_2 \in N$ , he should choose  $\mu_i^{j_1} > \mu_i^{j_2}$ . By varying the parameters  $\mu_i \in M_i, i \in N$  you can find those or other equilibrium, by solving the inequality (4). On the other hand, each

individually-optimal equilibrium is characterized by some set of parameters  $\mu_i \in M_i$ ,  $i \in N$ , and, accordingly, the preference on the set of payoff functions of all players.

It should also be noted that the parameters  $\mu_i^j$  can be given some game sense. Suppose the situation  $x^*$  is weak individually-optimal equilibrium. Then according to Theorem 1, it is the solution of inequalities (4), at least when  $\mu_i^j = \bar{\mu}_i^j$ ,  $i, j \in N$  (according to (3)). It is easy to see that for each fixed  $i \in N$  the value of parameter  $\mu_i^j$  indicates the desired value of the payoff function  $u_j(x)$ ,  $j \in N$  in compromise  $x^*$ , according to the player  $i \in N$ . Then and only then it will not profitable for player  $i \in N$  to deviate from situation  $x^*$ .

Theorem 1 also shows that the situation  $x^*$ , which is an individually-optimal equilibrium of the original problem (1) for fixed values of the vectors of parameters  $\mu_i \in M_i$ ,  $i \in N$ , that characterize the preferences of the players on the set of criteria, is stable for any player  $i \in N$  on the function  $v_i(x, \mu_i) = \min_{j \in N} (u_j(x) - S\mu_i^j) + S/n$ , that characterizes the utility of the situation  $x \in X$  comparatively to changing strategy  $x_i^*$  to any other. Thus, individually-optimal equilibrium  $x^*$  is a Nash equilibrium in the game, which has a normal form  $(X_i, v_j(x, \mu_i); i \in N)$ .

Questions arise, whether a player who is in individually-optimal equilibrium will be tempted to change its strategy as well as his preference on the set of payoff functions of other players? If so, can such a situation be considered the foundation of a stable agreement between the players? If not, then which of the individually-optimal equilibrium and under which players' preferences are stable in this sense?

**Definition 3.** Set, which consists of the situation  $\hat{x} \in X$  and the parameter vector  $\hat{\mu} = (\hat{\mu}_i)_{i \in N} \in M = \prod_{i \in N} M_i$

will be called the equilibrium in the preferences of players, if  $\forall x_i \in X_i \forall \mu_i = (\mu_i^j)_{j \in N} \in M_i \forall i \in N$   
 $\min_{j \in N} (u_j(\hat{x}) - S\hat{\mu}_i^j) \geq \min_{j \in N} (u_j(x_i, \hat{x}_{N \setminus i}) - S\mu_i^j)$ . Denote  $RE$  - set of equilibria in the preferences of players.

Thus, in equilibrium in the preferences of players  $(\hat{x}, \hat{\mu})$  it's not profitable to any player  $i \in N$  to change both its strategy  $\hat{x}_i$  and the parameter vector  $\hat{\mu}_i$ , which characterizes his preference on the set of payoff functions of other players.

**Theorem 2.** Set  $(\hat{x}, \hat{\mu})$  is equilibrium in preference of players if and only if situation  $\hat{x}$  satisfies the following inequalities:

$$\sum_{j \in N} u_j(\hat{x}) \geq \sum_{j \in N} u_j(x_i, \hat{x}_{N \setminus i}), \quad \forall x_i \in X_i, \quad i \in N. \quad (5)$$

If the set  $(\hat{x}, \hat{\mu})$  is an equilibrium in the preferences of players, the situation  $\hat{x}$  is weak individually optimal equilibrium, and the components  $\hat{\mu} = (\hat{\mu}_i)_{i \in N}$  are calculated according to (3).

1 \ 2	L <sub>2</sub>	H <sub>2</sub>
L <sub>1</sub>	A 8,5; 8,5	B -1,5; -1,5
H <sub>1</sub>	C 1,5; 1,5	D -0,5; -0,5

A particular equilibria that maximize the overall utility function of players can be identified among the set of equilibria in the preferences of players.

*Definition 4.* The equilibrium in the preferences of players  $(\hat{x}, \hat{\mu}) \in RE$  will be called optimal if it maximizes the overall utility function of players, i.e.  $(\hat{x}, \hat{\mu}) \in Arg \max \{ \xi(x) | (x, \mu) \in RE \}$ .

Conditions for optimality of equilibria in the preferences are set in the following theorem.

**Theorem 3.** Set  $(\hat{x}, \hat{\mu})$  is optimal equilibrium in the preferences of players if and only if situation  $\hat{x}$  satisfies the following conditions:  $\sum_{i \in N} u_i(\hat{x}) \geq \sum_{i \in N} u_i(x)$ ,  $\forall x \in X$ , and components of vector  $\hat{\mu} = (\hat{\mu}_i)_{i \in N}$  are calculated by

formula (3). Situation  $\hat{x}$ , which constitutes optimal equilibrium in the preferences of agents is Pareto-optimal situation of (1).

In particular, situation A will be optimal equilibrium in the preferences of agents for the game (Table 3). The vector of parameters that characterizes the preference of players, according to (3) will be the followed:  $\mu_1 = \mu_2 = (21 / 36, 15 / 36)$ . As a result, the situation A will become Nash equilibrium in game (Table 4), where the payoff of each player will be its utility function. Thus, taking in count the interests of each other will bring in situation A 9 units of additional profit to the first player and 6 to the second, in comparison with situation D (Table3). It should be noted that the "mechanisms" of taking in count the interests of one player by another are widely known and have long been used in practice (in particular, shares, securities, etc.). Therefore, this theory reveals just another aspect of their application for solving conflicts.

Note, that in [Voloshin, 2010] considered a somewhat different interpretation of the principle of "mutual consideration" of interests. It is offered to make the effective situation A (Table 3) a stable one (Nash equilibrium) by changing the initial matrix of the game in order to reallocate the players' payoffs. It requires a certain level of cooperation – at least an offer to "share gains". The proposed intuitive-understandable procedure requires a formal justification.

In this paper, constructive arrangements "taking into account the interests of each other" (on the basis of the above theory), which does not require co-operation of players, are proposed.

---

## Conclusion

The paper analyzes the principles of ineffective implementation of the Kyoto Protocol. Concepts which should be base for distribution of emission quotas mechanisms are proposed. It is alleged that the most appropriate principle of choosing the particular situation of collective decision in the allocation of emission rights is the principle of individual optimality, under which each party seeks to "compromise for the sake of conflict", choosing their strategies individually (noncooperative), but taking into account the interests of other participants.

---

## Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and Consortium FOI Bulgaria ([www.itea.org](http://www.itea.org), [www.foibg.com](http://www.foibg.com)).

---

**Bibliography**


---

- [Samuelson, 2008] Samuelson P., Nordhaus V. Economics.–M.: Williams, 2008.–1360c. (In Russian)
- [Voloshin, 2006] Voloshin O.F., Maschenko S.O. Decision Making Theory.- K.: VTC "Kyiv University", 2006.– 304 c. (In Ukrainian)
- [Voloshin, 2009] Voloshin A., Goritsyna I. Mechanisms of emission quotas distribution under the Kyoto Protocol ";. In: International Book Series" Information Science and Computing ", 2009.-V. 3 / 2009, N10. - P. 175 -181 (In Russian)
- [Maschenko, 2009] Maschenko S.O. Individually optimal equilibria of noncooperative games in the relations of preference // Cybernetics and Systems Analysis, 2009, № 1 .- P.171-179. (In Russian)
- [Maschenko, 2009] Maschenko S.O. Equilibria stable by preferences in noncooperative single-purpose games // Kyiv University Bulletin. Series: math and physics. Science. - 2009. - № 3. - S. 152-157. (In Ukrainian)
- [Harsanyi, 2009] Harsanyi J., Zelten R. General theory of equilibria choice in games.-S.-Pb.: Economic School, 2001.-206 pp. (In Russian)
- [Voloshin, 2009] A. Voloshin, I. Goritsyna, S. Maschenko. Methodological principles of quotas allocation for emissions of greenhouse gases; In: "Natural and Artificial Intelligence", ITHEA, Sofia, 2010. - P. 85-93. (In Russian)
- 

**Information about authors**


---



**Olexij Voloshin** – Professor, Taras Shevchenko National University, faculty of cybernetics. Kyiv, Ukraine. E-mail: [ovoloshin@unicyb.kiev.ua](mailto:ovoloshin@unicyb.kiev.ua)

*Scientific Interests: decision making, systems of decision making support, mathematical economics, expert systems, e-learning*



**Sergij Maschenko** – Associate Professor, Taras Shevchenko National University, faculty of cybernetics. Kyiv, Ukraine. E-mail: [msomail@yandex.ru](mailto:msomail@yandex.ru)

*Scientific Interests: optimization methods, decision making, systems of decision making support, game theory*

## НЕЧЕТКИЕ ОБОБЩЕНИЯ МОДЕЛИ РАСПРЕДЕЛЕНИЯ ЗАТРАТ

Алексей Ф. Волошин, Василий А. Лавер

**Abstract:** Рассматривается классическая задача распределения затрат, её интерпретация для случая распределения квот на выброс парниковых газов. Предлагаются нечеткие обобщения данной задачи, алгоритм её решения. Приводится иллюстративный пример.

**Keywords:** задача распределения затрат, задача распределения квот на выбросы парниковых газов, нечеткие модели и методы.

**ACM Classification Keywords:** I. Computing Methodologies – I.6. Simulation and modelling – I.6.5. Model Development – Modeling Methodologies.

### Введение

Классическая модель распределения затрат [Волошин, Мащенко, 2006] позволяет адекватно описать большое количество реальных процессов. Так, в [Волошин, Горицына, 2009] предлагается использовать её для распределения квот на выбросы парниковых газов (ПГ). Модель распределения затрат формулируется следующим образом:

Найти вектор  $x = (x_1, \dots, x_n)$  такой, что

$$\sum_{i=1}^n x_i = c \quad (1)$$

где  $n, n \geq 2$ , - количество агентов,  $c, c \geq 0$ , - общие затраты (общий лимит выброса ПГ),  $b_i$  - «потенциальный доход»  $i$ -го агента;  $x_i$  - доля затрат  $i$ -го агента (квота на выброс ПГ).

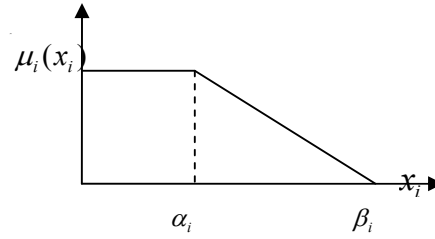
В связи с этим возникают две проблемы: 1. Как формировать «потенциальные доходы» каждого агента? 2. Какие механизмы распределения затрат использовать? При формировании «потенциального дохода» в [Волошин, Горицына, 2009] предлагается учитывать два эффекта влияния на загрязнение окружающей среды величину валового внутреннего продукта (ВВП) на душу населения и выбросы ПГ на 1 кв. км с учетом плотности населения; приводятся методики расчета величин  $b_i$ . Что касается механизмов распределения затрат, то предлагается априори задавать один из трех принципов распределения: выравнивание доходов, выравнивание затрат и пропорциональное распределение затрат. Основным недостатком «классической» модели является то, что в реальности исходные данные задаются неточно, нечетко, приближенно. Поэтому целесообразно рассмотреть нечеткие обобщения задачи распределения, в частности, задачи распределения квот на выбросы ПГ. В данной работе предлагаются три варианта нечетких обобщений рассматриваемой модели. Не уменьшая общности, рассмотрим принцип выравнивания затрат.

### Задача распределения квот при нечетких долях агентов

Имеется  $n$  агентов, которые упорядочены по возрастанию «потенциальных доходов»:  $b_1 \leq b_2 \leq \dots \leq b_n$ . Рассматривается задача распределения затрат в интерпретации распределения квот на выбросы ПГ. Общий лимит выброса равен  $c$  тонн  $\text{CO}_2$  - эквивалента ( $c > 0$ ). Квоты агентов  $x_i$  - нечеткие числа с функциями принадлежности  $\mu_i(x_i), i = \overline{1, n}$ .

Для всех  $i$  функция принадлежности агента  $i$  определяется по следующей формуле и имеет усеченно-трапецеидальный вид:

$$\mu_i(x_i) = \begin{cases} 1, & x_i \leq \alpha_i; \\ \frac{\beta_i - x_i}{\beta_i - \alpha_i}, & \alpha_i < x_i \leq \beta_i; \\ 0, & \beta_i < x_i, \end{cases}$$



$\alpha_i$  и  $\beta_i$  определяются в зависимости от величины  $b_i$ . Данный вид функций принадлежности можно интерпретировать следующим образом: пока агент  $i$  не превышает некоторый уровень выбросов  $\alpha_i$ , к нему не применяется никаких санкций и он обладает правом продавать свои квоты; промежуток между  $\alpha_i$  и  $\beta_i$  представляет значения выбросов, которые превышают квоту, но уровень превышения является приемлемым и задается значением  $r_i = 1 - \mu_i(x_i)$ . Если же значение выбросов ПГ превышает значение  $\beta_i$ , к агенту применяются санкции и штрафы.

Предположим, что агенты разделены на три группы в зависимости от уровня доходов - на «бедных» ( $i = \overline{1, n_1}$ ), «средняков» ( $i = \overline{n_1 + 1, n_2}$ ) и «богатых» ( $i = \overline{n_2 + 1, n}$ ), где  $1 \leq n_1 \leq n_2 \leq n$ .

Тогда

$$\alpha_i = \begin{cases} \gamma_1 b_i, & i = \overline{1, n_1} \\ \gamma_2 b_i, & i = \overline{n_1 + 1, n_2} \\ \gamma_3 b_i, & i = \overline{n_2 + 1, n} \end{cases}, \quad \beta_i = \begin{cases} \delta_1 b_i, & i = \overline{1, n_1} \\ \delta_2 b_i, & i = \overline{n_1 + 1, n_2} \\ \delta_3 b_i, & i = \overline{n_2 + 1, n} \end{cases},$$

где  $0 \leq \gamma_i \leq 1 (i = \overline{1, 3})$ ,  $\gamma_1 \leq \gamma_2 \leq \gamma_3$ ,  $0 \leq \delta_j \leq 1 (j = \overline{1, 3})$ ,  $\delta_1 \leq \delta_2 \leq \delta_3$  - некоторые коэффициенты.

Возникает вопрос, каким образом распределить квоты на выброс ПГ, учитывая при этом нечеткий вид  $x_i$ .

Решением задачи предлагается считать вектор  $(x_1, x_2, \dots, x_n)$ , который удовлетворяет следующим двум

условиям: 1)  $\sum_{i=1}^n x_i = c$ ; 2)  $\min\{\mu_1(x_1), \mu_2(x_2), \dots, \mu_n(x_n)\} \rightarrow \max$ .

Введем «индикатор» выполнения баланса – функцию  $v(x_1, x_2, \dots, x_n)$ :

$$v = 1, \sum_{i=1}^n x_i = c; v = 0, \sum_{i=1}^n x_i \neq c.$$

Функция принадлежности искомого вектора  $(x_1, x_2, \dots, x_n)$  будет определяться как:

$$\mu_c(x_1, x_2, \dots, x_n) = \min\{\mu_1(x_1), \mu_2(x_2), \dots, \mu_n(x_n), v(x_1, x_2, \dots, x_n)\}.$$

Решением задачи будем считать вектор, максимизирующий  $\mu_c(x_1, x_2, \dots, x_n)$ .

То есть для нахождения  $(x_1, x_2, \dots, x_n)$  нужно решить следующую задачу оптимизации:

$$\lambda \rightarrow \max, \mu_c(x_1, x_2, \dots, x_n) \geq \lambda.$$

Учитывая вид  $\mu_c(x_1, x_2, \dots, x_n)$ , получаем:

$$\lambda \rightarrow \max, \sum_{i=1}^n x_i = c, \mu_i(x_i) \geq \lambda, i = \overline{1, n}. \quad (2)$$



Так как  $\mu_1(x_1), \mu_2(x_2), \dots, \mu_n(x_n)$  линейны, задача (2) является задачей линейного программирования и может быть решена известными методами [Банди, 1988].

Рассмотрим числовой пример. Имеется два агента ( $n=2$ , для определенности - «Африка» и «Европа», см. [Волошин, Горицына, Мащенко, 2010], табл.3), между которыми необходимо распределить квоты на выброс ПГ величиной  $c=7$ . Функции принадлежности квот агентов задаются следующим образом (учитывая, что удельный вес выбросов «Европы» приблизительно в четыре раза выше аналогичного показателя для «Африки»):

$$\mu_1(x_1) = \begin{cases} 1, & x_1 \leq 1; \\ \frac{4-x_1}{3}, & 1 < x_1 \leq 4; \\ 0, & x_1 > 4. \end{cases} \quad \mu_2(x_2) = \begin{cases} 1, & x_2 \leq 4; \\ \frac{7-x_2}{3}, & 4 < x_2 \leq 7; \\ 0, & x_2 > 7. \end{cases}$$

$$\nu(x_1, x_2) = \begin{cases} 1, & x_1 + x_2 = 7; \\ 0, & x_1 + x_2 \neq 7. \end{cases}$$

Задача (2) примет следующий вид:  $\lambda \rightarrow \max, x_1 + x_2 = 7, \mu_1(x_1) \geq \lambda, \mu_2(x_2) \geq \lambda$ .

Рассмотрим возможные варианты изменений  $x_1$  и  $x_2$ .

При  $x_1 \leq 1, x_2 \leq 4$  не выполняется равенство  $x_1 + x_2 = 7$ ; при  $x_1 \leq 1, 4 \leq x_2 \leq 7$  и  $1 < x_1 \leq 4, x_2 \leq 4$  получим соответственно  $x_1 = 1, x_2 = 6$  и  $x_1 = 3, x_2 = 4$  при уровне  $\lambda = \frac{1}{3}$ .

Оставшийся вариант,  $1 \leq x_1 \leq 4, 4 \leq x_2 \leq 7$  порождает следующую задачу:

$$\lambda \rightarrow \max, x_1 + x_2 = 7; 1 \leq x_1 \leq 4; 4 \leq x_2 \leq 7; \frac{4-x_1}{3} \geq \lambda; \frac{7-x_2}{3} \geq \lambda. \quad (3)$$

Решением (3) будет тройка  $x_1 = 2, x_2 = 5, \lambda = \frac{2}{3}$ . Так как  $\lambda$  в этом случае достигает максимального значения, то это будет и решением всей задачи.

### Алгоритм решения задачи распределения затрат при нечетких долях агентов

Пользователь задает количество игроков –  $n$ , их доходы  $(b_1, b_2, \dots, b_n)$ , общий лимит выброса  $c$ , а также необходимые коэффициенты для всех трех групп доходности.

На основе полученных данных вычисляются функции принадлежности квот агентов.

Для каждой  $\mu_i(x_i), i = \overline{1, n}$ , возможны два существенных варианта: когда  $x_i \in [0, \alpha_i]$  или же когда  $x_i \in [\alpha_i, \beta_i]$ . Далее мы должны перебрать все возможные комбинации.

Для этого введем  $n$  – мерный булевый массив  $M$ . Изменяя значения массива с  $(0, 0, \dots, 0)$  до  $(1, 1, \dots, 1)$ , на каждом шагу будем решать следующую задачу:

$\lambda \rightarrow \max,$

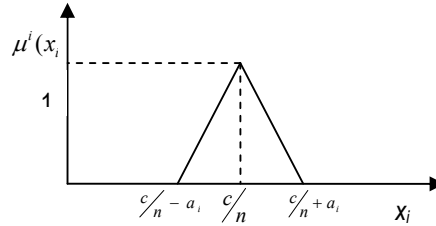
$$\sum_{i=1}^n x_i = c, \mu_i(x_i) \geq \lambda, x_i \in [d_i, e_i], \forall i = \overline{1, n}, \text{ где } d_i = \begin{cases} 0, & M_i = 1 \\ \alpha_i, & M_i = 0 \end{cases}, e_i = \begin{cases} \alpha_i, & M_i = 1 \\ \beta_i, & M_i = 0 \end{cases}.$$

За решение первоначальной задачи принимаем то из полученных решений, для которого  $\lambda$  будет максимальным.

### Задача распределения квот при нечетких долях агентов и заданном принципе распределения

Рассматривается принцип выравнивания квот. Каждому агенту соответствует доля в  $\frac{c}{n}$  тонн CO<sub>2</sub> эквивалента. Пусть  $a_i$  ( $i = \overline{1, n}$ ) – величина, на которую  $i$ -й агент имеет право отклониться от установленного распределения. Эта величина задается извне, она может зависеть от «достатка» агента. Тогда для  $i$ -го агента получаем следующий вид функции принадлежности допустимых отклонений от заданного распределения:

$$\mu^i(x_i) = \begin{cases} \frac{x_i/a_i - \frac{c/n - a_i}{a_i}}{\frac{c/n - a_i}{a_i}}, & x_i \in [c/n - a_i, c/n] \\ 1 - \frac{x_i - c/n}{a_i}, & x_i \in (c/n, c/n + a_i]; \\ 0, & \text{в других случаях.} \end{cases}$$



Формируется следующая функция принадлежности распределения  $x$  нечеткой цели:

$$\mu_G(x_1, x_2, \dots, x_n) = \min\{\mu^1(x_1), \mu^2(x_2), \dots, \mu^n(x_n)\}.$$

Эта функция является функцией принадлежности декартового произведения нечетких множеств допустимых отклонений от заданного распределения.

Функция принадлежности решения в этом случае будет иметь вид

$$\mu_D(x_1, x_2, \dots, x_n) = \min\{\mu_C(x_1, x_2, \dots, x_n), \mu_G(x_1, x_2, \dots, x_n)\}.$$

Решениями задачи будем считать альтернативы, максимизирующие  $\mu_D(x_1, x_2, \dots, x_n)$ . Алгоритм поиска решений аналогичен предыдущему, с той разницей, что необходимо учитывать изменения  $\mu^i(x_i)$ .

### Задача распределения квот при нечетких долях агентов, заданном принципе распределения и нечетком балансе выбросов

Предположим, что равенство  $\sum_{i=1}^n x_i = c$  выполняется нечетко:  $\sum_{i=1}^n x_i \cong c$ . Степень выполнения этого равенства определяется следующей функцией принадлежности

$$\nu(x_1, x_2, \dots, x_n): \quad \nu = \frac{\sum_{i=1}^n x_i - (c - \underline{c})}{\underline{c}}, \quad c - \underline{c} \leq \sum_{i=1}^n x_i \leq c; \quad \nu = \frac{\sum_{i=1}^n x_i - (c - \bar{c})}{\bar{c}}, \quad c < \sum_{i=1}^n x_i \leq c + \bar{c},$$

$\nu = 0$  иначе.

Тут  $\underline{c}, \bar{c}$  – значения допустимых отклонений  $\sum_{i=1}^n x_i$  от  $c$ .

Функция принадлежности распределения  $x$  ограничениям будет иметь вид:

$$\mu_C(x_1, x_2, \dots, x_n) = \min\{\mu_1(x_1), \mu_2(x_2), \dots, \mu_n(x_n), \nu(x_1, x_2, \dots, x_n)\}.$$

Функция принадлежности  $x$  цели –  $\mu_G(x_1, x_2, \dots, x_n) = \min\{\mu^1(x_1), \mu^2(x_2), \dots, \mu^n(x_n)\}.$

Функция принадлежности  $x$  решению –  $\mu_D(x_1, x_2, \dots, x_n) = \min\{\mu_C(x_1, x_2, \dots, x_n), \mu_G(x_1, x_2, \dots, x_n)\}.$

Решение задачи находится аналогично предыдущим случаям.

---

## Выводы

Рассмотренные нечеткие модели распределения затрат дают возможность более адекватно описывать реальные процессы, в частности, распределения квот на выброс парниковых газов. В будущем планируется усовершенствовать алгоритмы решения рассматриваемых задач на случай большой размерности, избежав использование процедур полного перебора при рассмотрении интервалов изменений  $x_i$ ; рассмотреть другие обобщения модели распределения затрат (в частности, распределение затрат при заданных функциях полезности агентов).

---

## Благодарности

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

---

## Библиография

- [Волошин, Мащенко, 2006] Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень - К.: ВТЦ "Київський університет", 2006.-304с.
- [Волошин, Горицына, 2009] Волошин А., Горицына И. Механизмы распределения квот на выбросы по Киотскому протоколу; In: International Book Series "Information Science and Computing", 2009. – V. 3/2009, N10. – P. 175-181
- [Банди, 1988] Банди Б. Методы оптимизации.. – М.: Радио и связь, 1988. -128с.
- [Волошин, Горицына, Мащенко, 2009] Волошин А., Горицына И., Мащенко С. Методологические принципы распределения квот на выбросы парниковых газов; In: "Natural and Artificial Intelligence", ITHEA, Sofia, 2010. - P. 85-93.

---

## Сведения об авторах



**Волошин Алексей** – доктор технических наук, профессор, Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: [ovoloshin@unicyb.kiev.ua](mailto:ovoloshin@unicyb.kiev.ua)

**Сфера научных интересов:** принятие решений, системы поддержки принятия решений, математическая экономика, экспертные системы, е-образование



**Лавер Василий** – Аспирант, Ужгородский национальный университет, математический факультет, Украина, 88000, Ужгород, ул. Университетская, 14, e-mail: [v.laver@gmail.com](mailto:v.laver@gmail.com)

**Сфера научных интересов:** принятие решений, системы поддержки принятия решений, математическая экономика, нечеткие модели и методы

## НЕЧЕТКАЯ МОДЕЛЬ УДОВЛЕТВОРИТЕЛЬНОГО РЕШЕНИЯ ЗАДАЧИ ВЫБОРА

Николай Н. Маляр

**Abstract:** Рассматриваются проблемы многокритериального выбора удовлетворительного решения при помощи аппарата нечеткой логики. Описаны методы построения и виды функции принадлежности. Приведен алгоритм моделирования нечеткого решения для данной задачи выбора.

**Keywords:** задача многокритериального выбора, нечеткая логика, модель удовлетворительного решения.

**ACM Classification Keywords:** H.1.1. Systems and Information Theory

### Введение

Важным компонентом исследования общественных явлений в разных направлениях человеческой деятельности выступает проблема принятия решения. Каким бы решение не было, будь то простое решение, или сложно организованный многоэтапный процесс, решение является актом выбора на множестве возможных вариантов (альтернатив). Выбор одних может быть значащим только для отдельного индивида; другие же, например, принимаемые в экономической сфере, могут существенно затрагивать интересы многих людей. Чем более богатым является множество альтернатив, тем выше вероятность получения наилучшего из возможных решений. Основаниями для возникновения ситуации выбора являются, с одной стороны, ограниченность ресурсов, а с другой — возможность их использования для удовлетворения различных потребностей. Таким образом, выбор — это совокупность действий, которые выполняет человек для достижения целей (удовлетворения потребностей) в условиях ограниченности. Каждая экономическая система сталкивается с необходимостью совершать те или иные виды выбора, связываемые с получением ответов на такие основные вопросы: что и сколько производить; кто, какую работу, как и в какие сроки должен выполнять; для кого предназначены результаты работы. Как показывает практика, реальные задачи выбора это задачи с многими критериями.

### Постановка задачи

В общем случае задачу многокритериального выбора можно сформулировать в следующем виде. Задано множество альтернатив решения некоторой проблемы -  $A = \{a_j\}_{j=1}^m$  и множество критериев  $K = \{k_i\}_{i=1}^n$  для оценки полезности альтернатив. Каждой альтернативе  $a_j$  лицо, принимающее решение (ЛПР) (или привлеченные к принятию решения эксперты) выставляют оценки  $x_{ij}$  по всему множеству критериев  $K$ . То есть  $x_{ij}$  это оценка  $j$ -й альтернативы по  $i$ -му критерию. Оценки выставляются или в баллах (например, всем привычная шкала оценок в системе образования) напрямую, или вычисляются в соответствующих единицах или в рамках нечеткой логики. Результаты оценивания представляются в виде матрицы решений.

Таблица 1. Матрица решений

	$a_1$	$a_2$	...	$a_m$
$k_1$	$x_{11}$	$x_{12}$	...	$x_{1m}$
$k_2$	$x_{21}$	$x_{22}$	...	$x_{2m}$
...	...	...	...	...
$k_n$	$x_{n1}$	$x_{n2}$	...	$x_{nm}$

Далее для каждого вектора оценок  $\bar{x}_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$  вычисляется, в зависимости от примененной модели многокритериального выбора, что соответствует адекватной ситуации принятия решения, функция полезности  $U_i$ . Наилучшей объявляется та альтернатива, для которой функция полезности имеет максимальное значение – в случае, например, максимизации прибыли, или минимальное значение, например, при минимизации убытков. При вычислении значений полезности могут приниматься во внимание веса критериев  $w_i$ .

Как показывает практика, поведению людей отвечают как оптимизационные шаги, так и шаги получения удовлетворяющих результатов. Подтверждению этого служат следующие примеры с экономической сферы приведенные Нобелевским лауреатом Герберт А. Саймоном [Саймон 1999]: предпринимателя может совсем не беспокоить максимальный доход, просто он хочет получать доход, какой считает для себя достаточным; в современных условиях известны случаи, когда акционеры-владельцы и реальные управленцы – это разные люди и последние могут быть не заинтересованы в максимизации прибыли; в условиях несовершенной конкуренции максимизация дохода – сомнительная цель, так как оптимальное решение для одной фирмы зависит от поведения других фирм.

Приведенные примеры позволяют высунуть гипотезу, что «достаточная прибыль» по содержанию, более психологически связана с уровнем достижения, как с максимизацией. Более того, условия наступления удовлетворенности не являются неизменяемыми, а определяются уровнями достижения, которые могут быть выше или более низкими в зависимости от жизненного опыта. На основе этой теории (подхода) целью фирмы является не максимизация, а достижение определенного уровня прибыли, удержание определенной доли рынка или определенного уровня продаж, то есть фирма будет пытаться достичь скорее всего удовлетворенности, чем максимизации.

Рассмотрим случай, когда множество альтернатив есть Паретовским множеством, а критерии эффективности измеряются в разных единицах измерения. Введем в рассмотрение фиктивную альтернативу  $T$ , которую назовем точкой «удовлетворения». Координаты этой точки представляют собой оценки по соответствующим критериям, значениями которых ЛПП может быть удовлетворено. Таким образом, точка  $T = (t_1, \dots, t_n)$  есть точкой пространства  $R_+$ . Эти оценки могут быть:

- а) точечными ( $t_i \in R_+$  - неотрицательные числа);
- б) интервальными ( $t_i = [\alpha_i, \beta_i] \in R_+$  - интервалы);
- в) нечеткими числами ( $t_i = \{t, \mu_i(t) \mid t \in R_+\}$  - замкнутые выпуклые подмножества на положительной полуоси  $R_+$ ). Последний случай включает также лингвистические оценки и два предыдущих случая.

Одной из проблем, связанных с оценкой удовлетворенности, является то, что сама удовлетворенность представляет собой нечеткое, размытое понятие, на значение которого сильное влияние оказывают суждения, восприятие и эмоции человека. В связи с этим предлагается использовать при измерении удовлетворенности лингвистические переменные, т.е. такие переменные, значениями которых являются не числа, а слова или предложения на естественном или формальном языке [Кофман, 1982].

Для значений лингвистических переменных, представляющих собой нечеткое подмножество, строится функция принадлежности, т.е. такая функция, которая каждому элементу из универсального множества всех возможных оценок ставит в соответствие число из интервала от 0 до 1, которое характеризует степень принадлежности данного элемента рассматриваемому нечеткому подмножеству.

**Определение.** Нечетким множеством  $B \subset Y$  называется совокупность пар вида  $(x, \mu_B(y))$ , где  $y \in Y$ , а  $\mu_B$  (функция  $Y \rightarrow [0;1]$ ) называется функцией принадлежности нечеткому множеству  $B$ . Значения  $\mu_B(y)$  для конкретного  $y$  называются степенью принадлежности этого элемента нечеткому множеству  $B$ .

### Методы построения и виды функции принадлежности

Известным есть тот факт, что качество принятого решения зависит от того насколько адекватно построенная функция принадлежности отражает знания эксперта или экспертов. В настоящее время можно выделить две группы методов построения функций принадлежности: прямые и косвенные методы [Борисов,1990].

*Прямые методы* характеризуются тем, что ЛПР задает правила определения значений функции принадлежности  $\mu_B(y)$ , характеризующей элемент  $y$  непосредственно. Значения его предпочтений на множестве элементов  $Y$  согласуются следующим образом :

1. Для любых  $y_1, y_2 \in Y$   $\mu_B(y_1) < \mu_B(y_2)$  тогда и только тогда, когда  $y_2$  предпочтительнее  $y_1$ , т.е. в большей степени характеризуется свойством  $B$ ;
2. Для любых  $y_1, y_2 \in Y$   $\mu_B(y_1) = \mu_B(y_2)$  тогда и только тогда, когда  $y_1$  и  $y_2$  безразличны относительно свойства  $B$ .

Прямые методы задания функции принадлежности, используются как правило, для измеримых понятий, таких как цена, прибыль, скорость, время, расстояние, давление, температура и т.д. Разновидностями прямых методов являются прямые групповые методы, когда, например, группе экспертов предъявляют конкретный объект, и каждый должен дать один из двух ответов: принадлежит или нет этот объект к заданному множеству. Число утвердительных ответов, деленное на общее число экспертов, дает значение функции принадлежности объекта к данному нечеткому множеству.

К прямым методам относятся также непосредственное задание функции принадлежности таблицей, графиком или формулой.

Как показывает анализ результатов исследований и решения практических задач, связанных с необходимостью обрабатывать информацию, что прямые методы в основном используются в качестве вспомогательных, т. к. характеризуются большой долей субъективизма.

*Косвенные методы* характеризуются тем, что значения функции принадлежности выбираются таким образом, чтобы удовлетворить заранее сформулированным условиям. В таких случаях экспертная информация является только исходной информацией для дальнейшей обработки. Дополнительные условия могут налагаться как на вид получаемой информации, так и на процедуру ее обработки. Эти методы, как правило, используются в тех случаях, когда нет элементарных измеримых свойств, через которые определяются нечеткие множества. К таким методам можно отнести статистический метод, метод парных сравнений, метод экспертных оценок и ряд других. Рассмотрим наиболее часто используемые методы.

Метод построения функции принадлежности на основе парных сравнений базируется на отражении мнения эксперта об относительной принадлежности элементов множеству или степени выраженности у них свойства, формализуемого множеством. Степень принадлежности элементов множеству определяется посредством парных сравнений. Мнения экспертов отражены в матрицах оценок.

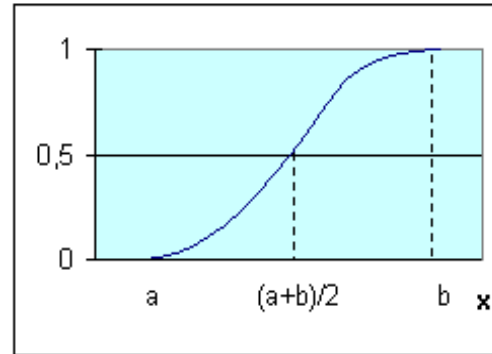
Метод статистических данных, основан на обработке статистической информации, может быть использован для формализации задачи выбора альтернатив. В качестве степени принадлежности элемента множеству принимается оценка частоты использования понятия, задаваемого нечетким множеством, для характеристики элемента. Функции принадлежности получаются при использовании специальных матриц подсказок [Борисов,1990]. Данный метод позволяет экспертам, как определять конкретное множество допустимых альтернатив, так и удаление не нужных альтернатив. В данном случае оценки отдельных экспертов можно рассматривать как независимые реализации случайной величины.

Построение функции принадлежности на основе экспертных оценок [Борисов, 1990]. Данный метод построения функций принадлежности базируется на использовании нечетких чисел, приблизительно равных некоторому четкому числу, и приближенных интервальных оценок, отражающих мнения экспертов

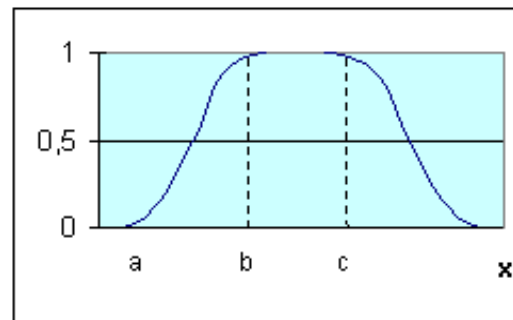
по рассматриваемому вопросу. Задача сводится к отысканию параметров заранее заданной функции, при решении которой используются результаты экспертного опроса. Как показал анализ различных источников, посвященных методам построения функций принадлежности, данный метод целесообразнее всего использовать при решении задач выработки и оценки альтернатив.

Основные виды функций принадлежности, применяемые в теории нечетких множеств [Алтунин,2000] приведены на рис.1.

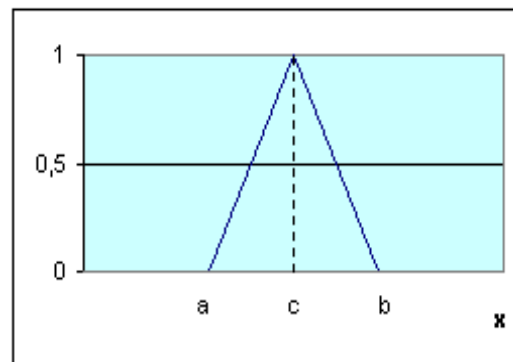
$$\mu_1(x, a, b) = \begin{cases} 0, & \text{если } x \leq a; \\ \frac{2(x-a)^2}{(b-a)^2}, & \text{если } a < x \leq \frac{a+b}{2}; \\ 1 - \frac{2(x-a)^2}{(b-a)^2}, & \text{если } \frac{a+b}{2} < x < b; \\ 1, & \text{если } x \geq b. \end{cases}$$



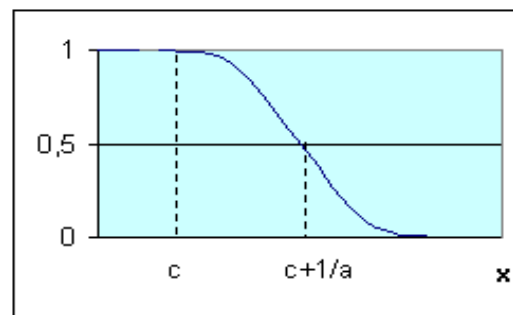
$$\mu_2(x, a, b, c) = \begin{cases} \mu_1(x, a, b), & \text{если } x < b; \\ 1, & \text{если } b \leq x \leq c; \\ 1 - \mu_1(x, c, c+b-a), & \text{если } x > c. \end{cases}$$



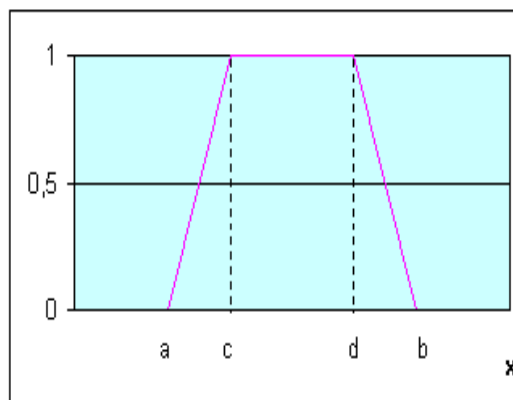
$$\mu_3(x, a, b, c) = \begin{cases} 0, & \text{если } x \leq a; \\ \frac{x-a}{c-a}, & \text{если } a < x \leq c; \\ \frac{b-x}{b-c}, & \text{если } c < x < b; \\ 0, & \text{если } x \geq b. \end{cases}$$



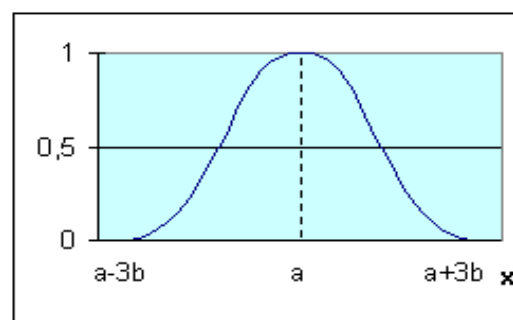
$$\mu_4(x, a, b, c) = \begin{cases} 1, & \text{если } x \leq c; \\ \left(1 + (a(x-c))^b\right)^{-1}, & \text{если } x > c. \end{cases}$$



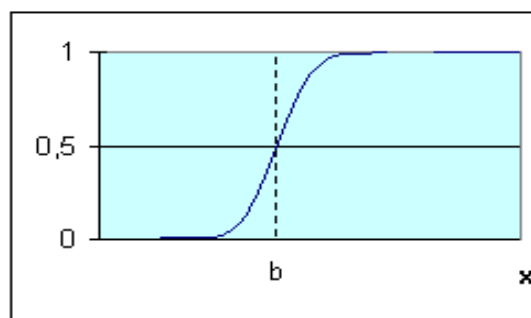
$$\mu_5(x, a, b, c, d) = \begin{cases} 0, & \text{если } x \leq a; \\ \frac{x-a}{c-a}, & \text{если } a < x < c; \\ 1, & \text{если } c \leq x \leq d; \\ \frac{b-x}{b-d}, & \text{если } d < x < b; \\ 0, & \text{если } x \geq b. \end{cases}$$



$$\mu_6(x, a, b) = \exp\left[-\frac{(x-a)^2}{2b^2}\right]$$



$$\mu_7(x, a, b) = [1 + \exp(-a(x-b))]^{-1}$$



**Рис 1.** Виды функций принадлежности.

### Алгоритм построения нечеткой модели решения

Предположим, что заданное множество альтернатив  $A = \{a_j\}_{j=1}^m$  может быть оценено при помощи совокупности критериев эффективности  $K = \{k_i\}_{i=1}^n$  и соответствующие оценки  $x_{ij}$  выражены в естественных единицах измерения для каждого из критериев. Ставится задача выбора наилучшей альтернативы. Пусть ЛПР может задать свою удовлетворенность по каждому из критериев, то есть свою точку “удовлетворения” и каким результатом относительно этой точки он был бы удовлетворен (термы для лингвистической переменной) [Маляр, 2005, 2006, 2008].

Построим, для данной ситуации модель решения используя теорию нечетких множеств в виде обобщенного алгоритма:

**Шаг 1.** Зададим для каждого критерия: точку “удовлетворения” в виде точечного результата или интервального или нечеткого числа и его полезность в виде весового коэффициента  $w_i$ .



**Шаг 2.** Выберем для лингвистической переменной «наилучшая» одно из значений терма: «близко», «не хуже», «намного лучше» и т.д.

**Шаг 3.** Построим нечеткую матрицу решений  $Z = (z_{ij})$  используя: матрицу решений  $X = (x_{ij})$ , точку «удовлетворения»  $T = (t_1, \dots, t_n)$  и значения лингвистической переменной.

**Шаг 4.** Произвести нормализацию весовых коэффициентов.

**Шаг 5.** Ищем решение задачи выбора в виде композиционного правила Заде ( $V = W \circ Z$ ).

**Примечание.** Для построения нечеткой матрицы решений (**шаг 3**), в зависимости от значений координат точки «удовлетворения» и терма лингвистической переменной могут использоваться различные виды функции принадлежности, приведенные на рис 1.

---

### Заключение

Данный подход позволяет представить модель задачи выбора через размытые множества. Используя его, ЛПР может строить свое размытое множество, которое зависит от множества критериев эффективности, точки «удовлетворения», его определения и способа построения функции принадлежности. Таким образом, нечеткое множество  $V$  будем считать моделью выбора, то есть его функцию принадлежности используем как функцию полезности альтернатив. Относительно значений этой функции может осуществляться ранжирование альтернатив.

---

### Библиография

- [Алтунин, 2000] Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях: Монография. Тюмень: Издательство Тюменского государственного университета, 2000, 352 с.
- [Борисов, 1990] Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решения на основе нечетких моделей: примеры использования. – Рига, "Знание", 1990, 184 с.
- [Кофман, 1982] Кофман А. Введение в теорию нечетких множеств. М: Радио и связь, 1982, 432с.
- [Маляр, 2005] Маляр М.М. Описання задач вибору на мові розмитих множин // Вісник Київського університету. Вип.4: Серія: фіз.-мат. науки, Київ, 2005.- С.197-201.
- [Маляр, 2006] Маляр М.М. Задача вибору та підхід до її розв'язання // Вісник СевДТУ. Вип.50: Інформатика, електроніка, зв'язок: Зб. наук. пр. – Севастополь: Вид-во СевДТУ, 2006.- С. 98-104.
- [Маляр, 2008] Маляр Н.Н. Применение нечеткой логики для задач коллективного выбора // Alexey Voloshin, Krassimir Markov, Krassimira Ivanova, Mykola Malyar, Iliia Mitov (ed.). Artificial Intelligence and Decision Making. International Book Series "INFORMATION SCIENCE & COMPUTING", Number 7, Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 2 / 2008, Institute of Information Theories and Applications FOI ITHEA - Sofia, Bulgaria, 2008. – P. 99 -102.
- [Саймон, 1999] Герберт А. Саймон. Теория принятия решений в экономической теории и науке о поведении // Вехи экономической мысли. Теория потребительского поведения и спроса. Т.1./ Под ред. В.М. Гальперина. СПб.: Экономическая школа, 1999г.

---

### Сведения об авторе

**Маляр Николай Николаевич** – декан математического факультета, заведующий кафедрой кибернетики и прикладной математики Ужгородского национального университета, кандидат технических наук, доцент, Украина, Ужгород, ул. Подгорная, 46; e-mail: malyartmm@gmail.com

## О ПРИНЯТИИ РЕШЕНИЙ ПРИ АНАЛИЗЕ МАЛЫХ ВОЗМУЩЕНИЙ ЛИНЕЙНЫХ МОДЕЛЕЙ

Всеволод Богаенко, Владимир Кудин

**Аннотация.** На основе вычислительного эксперимента разработан вариант системы поддержки принятия решений (СППР) с применением метода базисных матриц (МБМ) для анализа влияния малых возмущений в линейных моделях.

**Ключевые слова:** линейные модели, количественный и качественный анализ, малые возмущения, базисная матрица.

**ACM Classification Keywords:** I. Computing Methodologies – I.6. Simulation and modeling

---

### Введение

подавляющее большинство технических и экономических процессов описывается в классе линейных моделей (ЛМ). Например, основная макроэкономическая модель, модель Леонтьева (МЛ) [Волошин, 2004], описывается в виде системы линейных алгебраических уравнений (СЛАУ) при ограничениях на переменные в виде гиперпараллелепипеда. Учет неточности представления ЛМ (так называемая проблема адекватности математической и машинной модели) предопределяет разработку механизма согласования результатов проведения вычислений при разной точности представления данных модели. Некорректность модели при проведении вычислений в рамках ЛМ может существенно повлиять на качественные характеристики, например, величину ранга [Кудин, 2008]. Такие неточности зачастую обусловлены ограниченностью длины мантииссы при представлении чисел с плавающей запятой (*ошибки усечения, округления*) [Деммель 2001]. Использование разных типов данных при проведении вычислений (чисел с плавающей запятой одинарной, двойной, повышенной точности) существенно влияет на основные критерии оценки эффективности: быстродействие, точность и объемы вычислений.

---

### Цель исследования

Целью исследования являются: экспериментальный анализ влияния использования разных типов данных, алгоритмов, уровня обусловленности системы при построении вычислительных схем на основные параметры решения - точность самого решения и обращения матрицы, быстродействие; проверка эффективности вычислительных схем метода базисных матриц (МБМ) по указанным критериям на моделях заданной размерности; построение концепции принятия решения по достижению заданной эффективности вычислительного процесса.

---

### Концепция анализа

Концепция анализа ЛМ предполагается трехстадийной. Первая стадия содержит анализ типовой модели заданной размерности на разрешимость (существование, единственность решений СЛАУ), исследование свойств модели на основе базисного метода и алгоритма.

Вторая – проведение расчетов на основе вычислительных алгоритмов (без процедуры уточнения, с одно- или двухстадийной процедурой уточнения в каждом из них), при использовании разных типов данных (с плавающей запятой размерностью 64, 128, 256 бит) для СЛАУ с разными числами обусловленности. Фиксируются значения параметров быстродействия, точности решения и обращения матрицы, оценки объема используемой памяти.

Третья - построение функциональных зависимостей (интерполяционных многочленов) быстродействия, точности решения и обращения от используемых типов данных, алгоритмов, числа обусловленности.

## Постановка задачи

Предметом исследования является ЛМ:

$$Au = C, \quad (1)$$

где  $A = \{a_{ij}\}_{i,j=1,m}$  матрица размерности  $(n \times m)$ ,  $a_j = (a_{j1}, a_{j2}, \dots, a_{jm})$ ,  $j \in J = \{1, 2, \dots, n\}$ ,  $l = \{1, 2, \dots, m\}$  – строки матрицы  $A$ ,  $u = (u_1, u_2, \dots, u_m)^T$  – вектор переменных,  $C = (c_1, c_2, \dots, c_n)^T$  – вектор градиента ограничений модели.

Для решения и анализа (1) в МБМ [Кудин, 2002] введены в рассмотрение строчные базисные матрицы. Базисные матрицы в ходе итераций последовательно изменяются замещением строк вспомогательной СЛАУ строками (нормальями ограничений) основной СЛАУ. В общем случае в исследуемой модели количество ограничений превышает количество переменных.

В работе исследовано влияние использования разных вариантов организации вычислений на эффективность алгоритмов метода базисных матриц. Предложена концепция построения системы поддержки принятия решения по организации вычислений на линейных моделях для достижения заданных значений параметров по основным критериям: точности и быстродействию.

## О влиянии малых возмущений в “малых” ЛМ методом базисных матриц (МБМ)

Структурно исследование было построено как многошаговая алгоритмическая процедура: построение последовательности модельных задач малой размерности со свойством плохой обусловленности с некоторого шага; нахождение на основе метода базисных матриц основных элементов МБМ; вычисление чисел обусловленности; графическое представление основных функциональных зависимостей. Эксперимент демонстрирует количественно-качественные связи основных параметров исследуемой модели и метода базисных матриц.

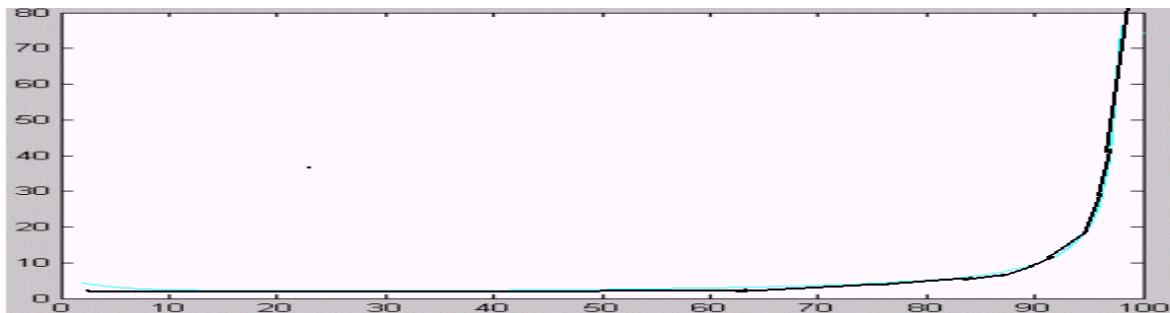


Рис. 1. Зависимость значений норм ведущих столбцов от номера итерации

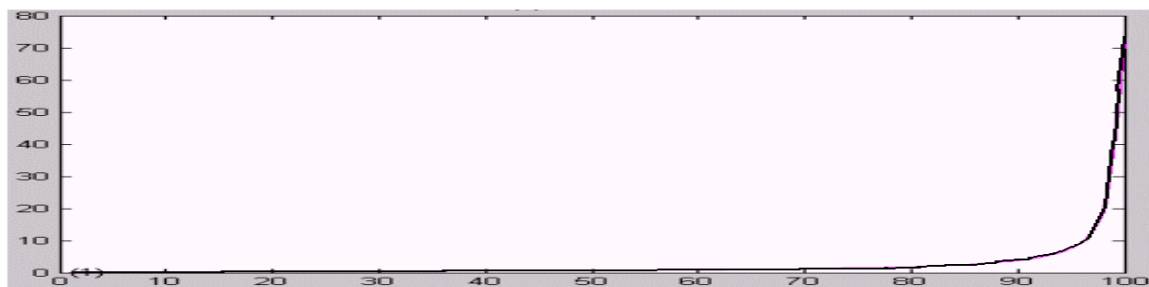
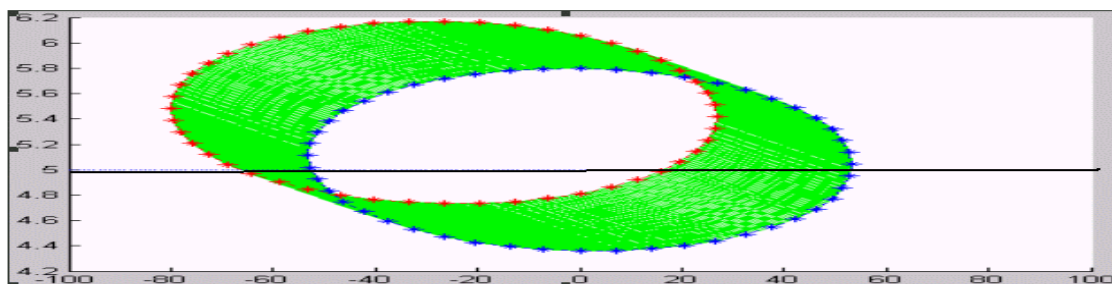


Рис. 2. Зависимость значений чисел обусловленности от номера итерации



Рис. 3. Зависимость значений ведущих элементов от номера итерации

Рис. 4. Эллипсы принадлежности решений на итерациях метода  
(оси координат компоненты вектора решения)

Как видно из графиков, вхождение СЛАУ в состояние плохой обусловленности характеризуется количественными значениями (числа обусловленности, ведущих элементов, норм ведущих столбцов обратных матриц): число обусловленности резко увеличивается; ведущие элементы симплексных преобразований МБМ становятся как угодно близкими к нулю; нормы столбцов увеличиваются в состоянии плохой обусловленности; в состоянии плохой обусловленности малым возмущением в модели отвечают значительные качественные отклонения "соседних решений", "вход" системы в состояние плохой обусловленности предопределяет „вытягивание" эллипсоподобной фигуры относительно одной из осей (рис. 4.).

### Вычислительный эксперимент анализа малых возмущений в ЛМ «средней» размерности

Быстродействие алгоритмов зависит только от размерности матрицы. Для вычислительного эксперимента были выбраны три процедуры МБМ: без уточнения решения (0), с одностадийным (1) и двухстадийным (2) уточнением.

Возможность применения разных типов данных: числа с плавающей запятой двойной точности (Double), 128-битные числа (Dd) и 256-битные числа (Qd) при разных модификациях алгоритмов МБМ позволяет построить систему поддержки принятия решения, которая на основе заданных пользователем ограничений на такие параметры алгоритмов, как быстродействие и точность, позволит выбрать наилучший по этим критериям алгоритм.

Выбор базируется на эвристических зависимостях параметров алгоритма и значений критериев от параметров матрицы ограничений  $n$  и некоторого параметра быстродействия системы для заданного типа данных и может быть оценен как:

$$T^2(n) = (3n^3 + n^2)t_c, \quad (2)$$

где значение параметра  $t_c$  должно оцениваться для каждой конкретной аппаратно-программной платформы экспериментальным путем.

Стоит отметить, что на одной и той же вычислительной платформе, отношение быстродействия алгоритмов, использующих разные типы данных, будет постоянным. Данные вычислительных экспериментов показали, что на выбранной для тестирования платформе, алгоритм, использующий 128-битные числа, был в ~35 раз медленней, чем алгоритм, использующий 64-битные числа, а в случае 256-битных чисел – в ~450 раз медленней. Такое существенное замедление объясняется тем, что операции с числами с плавающей запятой большого размера не реализованы аппаратно, в отличие от операций над 64-битными числами.

Что касается объема используемой алгоритмами памяти, то он пропорционален размеру используемого типа данных и будет, по сравнению с 64-битными числами, вдвое большим для 128-битных и вчетверо большим для 256-битных.

Для построения эвристических зависимостей между точностью решения и числом обусловленности была проведена серия вычислительных экспериментов. Используемая для тестирования аппаратная платформа – процессор AMD Athlon64 с реальной тактовой частотой 1.8Ghz, 512Mb оперативной памяти.

Проводилось решение разными алгоритмами МБМ СЛАУ размерностью 256x256.

В качестве критерия точности брались точность машинного решения  $u_o(1)$  в сравнении с аналитическим (точным)  $u = (1, 0, \dots, 0)$ ,  $\varepsilon_2 = \|u_o - (1, 0, \dots, 0)\|$ . Полученные экспериментальные данные приведены в табл. 1.

На основе этих данных (и данных при других размерностях) построены зависимости точности решения от числа обусловленности (Рис. 1.), в котором используются следующие обозначения: double – числа с плавающей запятой двойной точности (64битной размерности), dd – числа с плавающей запятой размерности 128 бит, qd – числа с плавающей запятой размерности 256 бит, +0( ), +1, +2 – количество итераций уточнения,  $\log_{10} c$  – порядок числа обусловленности системы,  $-\log_{10} \varepsilon_2$  – порядок точности решения.

Таблица 1. Зависимость порядка точности решения  $\varepsilon_2(\log_{10} \varepsilon_2)$  от числа обусловленности ( $m = 256$ )

$\log_{10} c$	6	7	8	9	10	11	12	12
Без уточнения решения (0)								
double	-15.81	-12.29	-8.89	-2.75				
dd	-46.16	-42.45	-39.07	-34.66	-32.50	-27.71	-22.33	-19.33
qd	-113.35	-107.94	-104.76	-100.43	-96.25	-93.69	-86.65	-85.01
Одностадийное уточнение (1)								
double+1	-12.40	-10.55	-5.37	-2.70				
dd+1	-45.92	-43.43	-36.86	-31.96	-29.12	-26.33	-22.43	-17.85
qd+1	-109.58	-108.57	-103.25	-95.93	-93.22	-91.00	-87.99	-80.48
Двух стадийное уточнение (2)								
double+2	-10.85	-8.55	-6.75	-2.27				
dd+2	-44.41	-42.15	-36.89	-32.16	-31.15	-26.44	-21.90	-20.09
qd+2	-101.21	-106.73	-107.70	-100.30	-93.89	-87.59	-86.52	-85.47

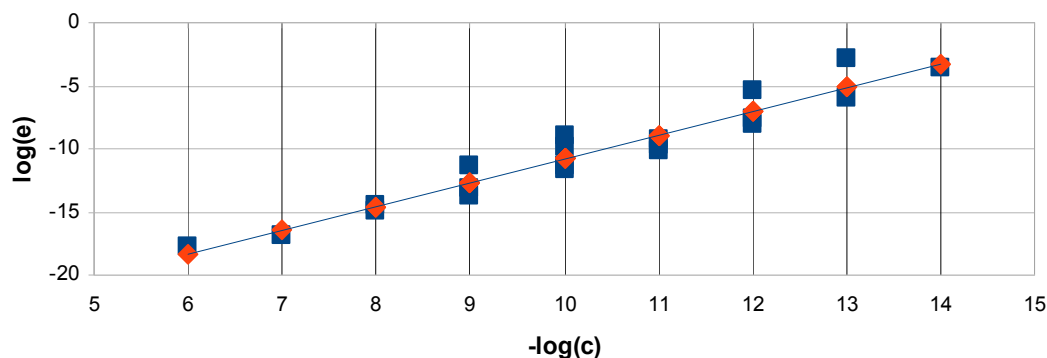


Рис. 1. Зависимость порядка точности  $\varepsilon_2$  ( $\log_{10} \varepsilon_2$ ) от числа обусловленности

**Вычислительный эксперимент установил близкие к линейным зависимости:** точности решения (и обращения матрицы) от числа обусловленности (при фиксированном типе данных, алгоритме) и размерности модели, точности решения от типа данных (при фиксированном числе обусловленности и алгоритме), *близкие к полиномиальным зависимости* быстродействия от размерности задачи (при фиксированных типах данных, числе обусловленности и размерности модели).

Учитывая полученные зависимости, алгоритм “работы” системы поддержки принятия решений можно описать следующим образом.

**Входные данные:**  $t_m$  - максимально допустимое время расчётов;  $\varepsilon_m$  - максимально допустимая погрешность решения;  $P$  - приоритетность критериев (времени и погрешности).

1) Построение множества алгоритмов, которые удовлетворяют условию относительно времени расчётов:

$A_t = \{A_i : T_i(n) < t_m\}$ , где  $T_i(n)$  - время работы алгоритма  $A_i$ ,  $i = \overline{1,6}$ , которое оценивается по формуле (2) (в общем случае на множестве алгоритмов).

2) Построение множества алгоритмов, которые удовлетворяют условию относительно погрешности:

$A_\varepsilon = \{A_i : E_i(n, \alpha) < \varepsilon_m\}$ , где  $E_i(n, \alpha)$  - оценка погрешности обращения матрицы или решения СЛАУ для алгоритма  $A_i$ ,  $i = \overline{1,6}$ , в общем случае из множества алгоритмов.

3) Если  $A_t \cap A_\varepsilon \neq \emptyset$ , в зависимости от приоритетности выбирается самый быстрый (или самый точный) алгоритм из множества приемлемых значений критериев  $A_t \cap A_\varepsilon$ .

4) В случае, когда  $A_t \cap A_\varepsilon = \emptyset$ ,  $A_t \neq \emptyset$  ( $A_\varepsilon \neq \emptyset$ ), в зависимости от приоритетности выбирается самый быстрый алгоритм из  $A_\varepsilon$  (или самый точный алгоритм из  $A_t$ ).

5) Если  $A_t = \emptyset$  или  $A_\varepsilon = \emptyset$  в зависимости от приоритетности выбирается самый быстрый или самый точный алгоритм из  $A_t \cup A_\varepsilon$  (лучший из неприемлемых).

## Выводы

Построенная концепция принятия решения на основе МБМ обладает следующими возможностями:

- анализировать и находить решения (и уточнять при необходимости) линейных систем уравнений за фиксированное время с заданной точностью по системе критериев;
- использовать решения исходной задачи для типовых моделей (при разных вариантах организации вычислительного процесса) для анализа и построения механизма принятия решений;

---

- строить экстраполяционные зависимости (для точности решений и быстродействия алгоритмов) от размерности (типовой модели) при фиксированном значении числа обусловленности и типе данных.

---

### Благодарности

---

Авторы благодарны проф. Волошину А.Ф. за консультации при написании статьи.

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

---

### Литература

---

[Волошин, 2004] Волошин О.Ф. Методи аналізу статичних балансових еколого-економічних моделей великої розмірності // Наукові записки, Т. VII, КПДВ "Педагогіка", Київ, 2004,- С. 43-55

[Кудин. 2008] Кудин В., Кудин Г., Волошин А. Анализ свойств модели Леонтьева при нечетко заданных параметрах с применением метода базисных матриц // Information Science & Computing, International Book Series, Number 7, Volume 7, ITHEA, SOFIA, p. 86-90, 2008.

[Деммель 2001] Деммель Дж. Вычислительная линейная алгебра. Теория и приложение М.: Мир, -2001.-430с.

---

### Информация об авторах

---

**Всеволод О. Богаенко** – к.т.н., с.н.с., Киев, Институт кибернетики имени В.М. Глушкова НАН Украины, Украина; e-mail: [sevab@ukrnet.ua](mailto:sevab@ukrnet.ua)

**Владимир И. Кудин** – доктор технических наук, Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: [V.I.Kudin@mail.ru](mailto:V.I.Kudin@mail.ru)

## ИССЛЕДОВАНИЕ МОДИФИЦИРОВАННОЙ СИСТЕМЫ МАРТИНГЕЙЛ ДЛЯ ТРЕЙДИНГА НА ВАЛЮТНЫХ РЫНКАХ

Юрий Зайченко, Владислав Плаксин

**Abstract:** *Decision Support System – modified Martingale is suggested in this paper. This DSS differs from classical Martingale by using preliminary training for finding optimal system parameters and financial indicators like MA (Moving Average) for estimate financial process trend and reveal its change. The corresponding software of the suggested DSS was elaborated and experimental investigations were carried out. This system was applied at currency market FOREX and used for determining the optimal trading strategy. The experimental results of its application are presented*

**Keywords:** *trading strategy at market, FOREX, decision support system, financial indicators, modified Martingale.*

**ACM Classification Keywords:** *H.4 INFORMATION SYSTEMS APPLICATIONS - H.4.2. Types of Systems Decision Support*

---

### Введение

Проблема организации трейдинга на валютных рынках представляет значительный теоретический и практический интерес. Для выбора оптимальной стратегии инвестора и принятия соответствующих решений необходимо осуществлять прогнозирование финансовых процессов на рынках, в частности, котировок валют (кросс-курсов). Сложность решения этой задачи заключается в том, что финансовые процессы носят стохастический характер и зависят от множества факторов, значения которых неизвестны. Для эффективной работы инвесторов и трейдеров на валютных рынках разработан ряд СППР-советников трейдеров, цель которых обеспечить поддержку принимаемых решений по купле-продаже валют [1-4].

К числу наиболее известных подобных систем относится система Мартингейл [5]. Эта система имеет ряд недостатков, а именно:

1. Для инвестора (трейдера) необходим неограниченный стартовый капитал;
2. Отсутствие гибкости, система не настраивается под специфику конкретных кросс- курсов валют.

Поэтому в работе [6] был сделан вывод о нецелесообразности использования этой системы в классическом варианте и предложена ее модификация- МБ АС Мартингейл, состоящая в следующем:

1. По сигналу финансового индикатора открывать две встречных сделки “buy” и “sell” во время любого движения рынка (вверх и вниз);
2. Уменьшить входящие лоты с целью уменьшения риска от возможных просадок.

Предварительные тесты по практическому применению СППР МБАС Мартингейл на рынке FOREX показали ее эффективность и подтвердили преимущества над классической системой Мартингейл. Вместе с тем для повышения эффективности ее работы требуется ее интеграция с различными финансовыми индикаторами для прогнозирования будущего состояния рынка, в частности, сохранится ли текущая тенденция рынка (рост или падение) или она изменится.

Для этой цели можно использовать различные индикаторы, такие, как скользящее среднее (MA), взвешенное скользящее среднее (WMA), экспоненциальное сглаживание (EMA) и др [7].

Целью настоящей работы является исследование целесообразности использования финансовых индикаторов типа MA, WMA в составе СППР Мартингейл, а также ее обучения для поиска оптимальных параметров, таких как ТП (тейк-профит) и СЛ (стоп-лосс), определяющих моменты окончания сделок по величине достигнутой прибыли (ТП) или потерь (СЛ).



---

### Описание алгоритма работы СППР МБАС Мартингейл

---

Весь цикл работы системы МБАС Мартингейл разбивается на два этапа: обучение и практическое использование.

На этапе обучения происходит поиск оптимальных параметров системы, а именно, величин ТП, СЛ, а также периодов усреднения для сделок типа buy и типа sell. Для этого используются алгоритмы обучения по методам сопряженных градиентов и стохастической аппроксимации. В дальнейшем для этих целей предполагается использование нейронных сетей.

После окончания этапа обучения переходим к этапу использования системы. На этом этапе реализуется алгоритм работы модифицированной МБАС Мартингейл. Открываются одновременно две сделки: одна со стратегией buy, а другая со стратегией sell.

Особенность системы Мартингейл состоит в том, что после каждого неудачного результата объем заказа удваивается. Каждая из сделок продолжается до момента остановки по одному из следующих правил:

- достижения максимальной прибыли ТП;
- достижения максимальных потерь СЛ.

Кроме того, если изменение скользящего среднего (дельта MA) меняет знак, то это признак начала изменения тенденции рынка и сигнал инвестору (трейдеру) изменить свою стратегию, например, закрыть текущую сделку и открыть новую противоположного типа.

---

### Описание результатов экспериментальных исследований

---

Были проведены эксперименты по использованию МБАС Мартингейл в комбинации с индикатором MA при торговле на валютном рынке FOREX

на интервале 07.2009- 03.2010. на кросс- курсах валют euro/USD.

При этом вводилось ограничение на сделку: пока текущая сделка не закроется, новая такого же типа не будет открыта. Сделки «buy» и «sell» работают независимо.

В результате исследований были определены следующие параметры настройки системы: начальный депозит 50000 USD , стартовый лот 0,1.

Были проведены 3 варианта экспериментов, отличающиеся значениями параметров ТП и СЛ, а именно:

#### Вариант 1

ТП= 160, СЛ= 60, период для MA1= 30 ( сделка типа «buy»), период MA2= 48 ( сделка типа «sell»).

#### Вариант 2

ТП=140, СЛ= 80, период MA1 = 12, период MA2 = 21.

#### Вариант 3

ТП= 160, СЛ 120 , период MA1= 3, период MA2= 54.

Основные результаты применения СППР МБАС Мартингейл с указанными настройками приводятся в таблице 1.

Анализ результатов таблицы 1 позволяет сделать следующие выводы.

1. Наибольшая чистая прибыль ( 173058 USD) достигается в варианте1, наименьшая – в варианте 3.
2. Наибольший профит- фактор (2,34) в варианте 1 , наименьший – в варианте 3 (1,59).
3. Наибольшая максимальная просадка – в варианте 1 ( 50021 USD) а наименьшая – в варианте 3 (4322 USD ).
4. Наибольшие потери – в варианте 1, а наименьшие – в варианте 3.
5. Таким образом, вариант 1 является наиболее рисковым и в то же время наиболее прибыльным, а вариант 3 – наоборот, наименее рисковым и наименее прибыльным.

Следовательно, стратегия по варианту 1 подходит для наиболее рискованных инвесторов, а по варианту 3 - для наиболее осторожных инвесторов, не склонных к авантюрам.

На рис 1-3 приведены результаты изменения доходов инвесторов для вариантов 1-3 соответственно.

**Таблица 1.** Результаты экспериментальных исследований СППР МБАС Мартингейл

№п/п	Показатели	Вариант1	Вариант2	Вариант3
1	Начальный депозит, USD	50000	50000	20000
2	Общий доход, USD	302050,13	141154,88	29843
3	Общие потери, USD	128991,36	-79943,62	-18789
4	Чистая прибыль, USD	173058,77	61211,26	11054,45
5	Профит-фактор общий доход/общие потери	2,34	1,77	1,59
6	Максимальная просадка, USD	50021	23375,65	4322,5
7	Относительная просадка, %	33,45	35,59	2161,25
8	Наибольшая прибыль, USD	80860,16	35547,84	5118
9	Наибольшие потери, USD	-15872	-10247	-1959,84
10	средняя прибыль, USD	8883,83	2663	806,59
11	Средние потери, USD	-1121	-768,69	-361,33
12	Максимальное число последовательных выигрышей	3	2	3
13	Максимальное число последовательных потерь	12	10	4
14	Среднее число последовательных выигрышей	1	1	1
15	Среднее число последовательных потерь	4	2	2

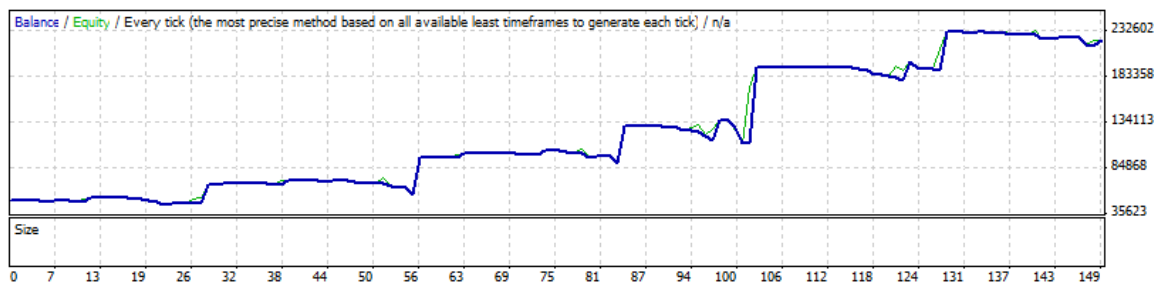


Рис. 1. Динамика изменения дохода инвестора для варианта 1

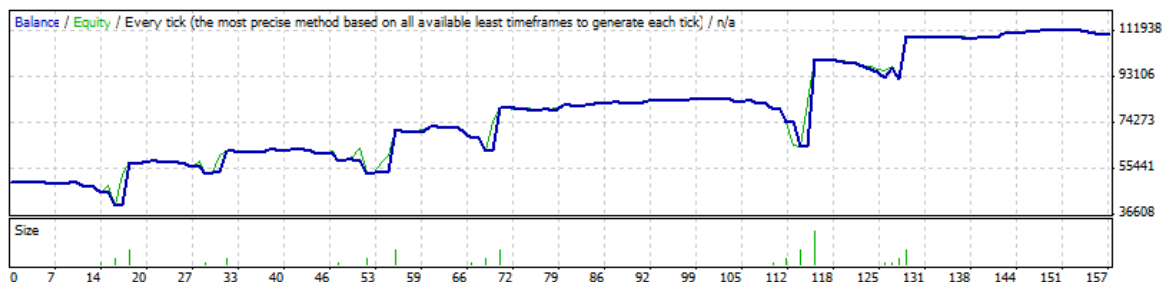


Рис. 2. Динамика изменения дохода инвестора для варианта 2

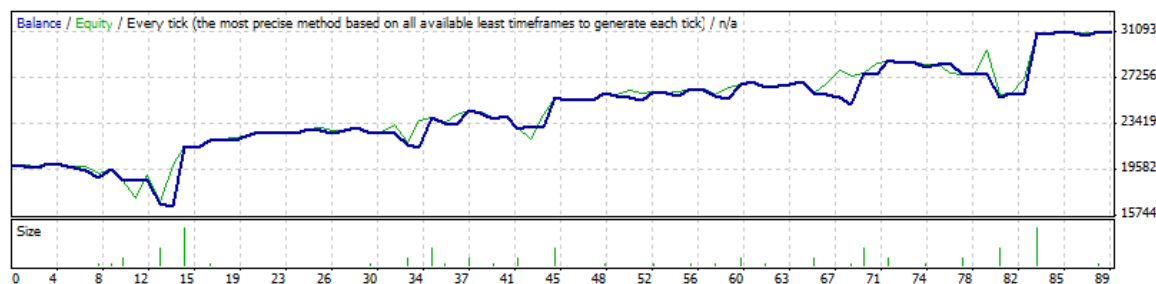


Рис.3. Динамика изменения дохода инвестора для варианта 3

## Выводы

1. В работе предложена модифицированная система Мартингейл - СППР для трейдинга на валютных рынках.
2. СППР использует предварительное обучение для поиска оптимальных параметров настройки системы, а также финансовый индикатор в виде скользящего среднего (МА) для отслеживания поведения рынка и изменений стратегии.
3. Проведены экспериментальные исследования разработанной системы на рынке FOREX, которые позволяют оценить ее эффективность в практической работе в процессе трейдинга.

## Благодарности

"The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria [www.ithea.org](http://www.ithea.org), and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine [www.aduis.com.ua](http://www.aduis.com.ua)".

## Литература

- Джеффри Оуэн Кац, Донна Л.Маккормик.Энциклопедия торговых стратегий. М.:Альпина Бизнес Букс, 2006, 400 с
- Шарп, У.Ф., Александер Г.Дж., Бэйли, Дж. В. (1997). Инвестиции.-М.ИНФРА-М,2006.1028 с.
- Брейли Ричард, Майерс Стюарт. Принципы корпоративных финансов. М.: «Олимп Бизнес», 2007,1008 .
- Дамодаран Асват. Инвестиционная оценка. Инструменты и методы оценки любых активов. –М.:Альпина Бизнес Букс, 2007, 1340 с.
- Энциклопедия финансового риск-менеджмента. Под редакцией А.Лобанова, А.Чугунова. М.: Альпина Бизнес Букс, 2009, 932 с.
- В. Плаксин. Модернизация системы Мартингейл и анализ ее работы на валютных рынках.-XIV International conference "Knowledge- Dialogue- Solution", KDS2-2009/.- Kiev, September 2009.-pp. 112-118.
- Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.:Мир, 1974.

## Информация об авторах

**Зайченко Юрий Петрович**, профессор, д.т.н., декан факультета «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14.

тел: 38044-406-83-93, e-mail: [ZaychenkoYuri@ukr.net](mailto:ZaychenkoYuri@ukr.net) , [baskervil@voliacable.com](mailto:baskervil@voliacable.com)

**Плаксин Владислав Константинович**, аспирант «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14.

Тел +380678722222, e-mail : [v.plaksin@forex4you.com.ua](mailto:v.plaksin@forex4you.com.ua)

## ПРОГНОЗИРОВАНИЕ РИСКА БАНКРОТСТВА БАНКОВ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ С ПРИМЕНЕНИЕМ НЕЧЕТКИХ НЕЙРОННЫХ СЕТЕЙ

Юрий Зайченко, Ольга Войтенко

**Abstract:** The problem of banks bankruptcy risk forecasting is considered. The application of fuzzy neural networks for this problem solution, ANFIS and TSK was suggested. The experimental investigations of application of FNN for bankruptcy forecast were carried out and comparison with classical methods was performed. The experimental investigations have proved the high efficiency of fuzzy neural networks for bankruptcy risk forecasting and their advantages over classical methods.

**Keywords:** bank bankruptcy risk forecasting, fuzzy neural networks.

**ACM Classification Keywords:** H.4 INFORMATION SYSTEMS APPLICATIONS - H.4.2. Types of Systems Decision Support

---

### Введение

В современной экономике проблема банкротства играет исключительно важную роль. Поскольку банковская система является фундаментом эффективно развивающейся экономики, своевременное определение риска банкротства банков является исключительно важной.

Экономический кризис конца 2008-2009 годов показал, что банковская система Украины крайне неустойчива. Показатели банков резко ухудшились в результате девальвации гривны (на 70%) в 2008г.) и резкого экономического спада. С начала кризиса из банков изъято более 80 млрд. гривен. Это привело к массовым неплатежам, а многие кредитные учреждения находятся на грани дефолта по взятым обязательствам. За полтора года более, чем в 20 банках введены временные администрации, около 10 банков ликвидированы либо их активы переведены в другие банки.

Проблема эффективного прогнозирования банка-банкрота является крайне актуальной как на макроуровне, так и на микроуровне при выборе рядовым гражданином банка для сохранения своих вложений. На данный момент существует множество методик определения рейтинга банков (Кромонова, WebMoney, CAMEL, методика агентства Moody's S&P и др.) [1,2]. Однако общим их недостатком является то, что они работают при полных и достоверных данных и потому не всегда дают правильные результаты, а порой использование разных методик приводит к противоречивым результатам. В связи с вышеуказанным, является целесообразным разработка новой методики прогнозирования банкротства банков в условиях неопределенности и неполноты исходной информации.

Целью настоящей работы является разработка нового подхода к прогнозированию риска банкротства банков на основе применения нечетких нейронных сетей (ННС) TSK и ANFIS [3] и сравнение полученных результатов с четкими методами (Кромонова) и методикой, разработанной в Белорусском государственном экономическом университете.

---

### Экспериментальные исследования

В ходе выполнения работы были разработаны программы, реализующие ННС TSK и ANFIS в среде Borland C++ Builder 4. Были собраны статистические данные по 170 коммерческим банкам Украины [1] и проведены широкие экспериментальные исследования эффективности прогнозирования банкротства банков различными методами и методиками. Некоторые из полученных результатов приводятся ниже.

Первая группа экспериментов использовала данные за 01.01.2008 (фактически за два года до потенциального банкротства).

#### Эксперименты 1-2.

Обучающая выборка = 120, проверочная += 50 банков. Количество правил = 5.

Входные переменные: Assets, capital, Cash, (Liquid assets), Household deposits, Liabilities. Соответствующие результаты приводятся в таблице 1.

**Таблица 1**

Результаты	TSK	ANFIS
Общее количество ошибок	5	6
%% ОШИБОК	10%	12%
ошибок первого рода	0	0
ошибок второго рода	5	6
Размер проверочной выборки	50	50

Обучающая выборка = 120, проверочная += 50 банков. Количество правил = 10.

Входные переменные: Assets, capital, Cash, (Liquid assets), Household deposits, Liabilities. Результаты приводятся в табл.2

**Таблица 2**

Результаты	TSK	ANFIS
Общее количество ошибок	6	7
%% ОШИБОК	12%	14%
ошибок первого рода	1	1
ошибок второго рода	5	6
Размер проверочной выборки	50	50

В следующей серии экспериментов выполнялся поиск оптимальных входных переменных. Период тот же.

### Эксперимент 3.

Обучающая выборка = 100, проверочная = 70. Количество правил = 5

Входные переменные: прибыль текущего года, чистый процентный доход, чистый комиссионный доход, чистые затраты на формирование резервов, чистая прибыль/убыток банка. Результаты экспериментов приведены в таблице 3.

**Таблица 3**

Результаты	TSK
Общее количество ошибок	13
%% ОШИБОК	19%
ошибок первого рода	6
ошибок второго рода	7
Размер проверочной выборки	70

**Таблица 4**

Результаты	TSK
Общее количество ошибок	7
%% ОШИБОК	10%
ошибок первого рода	0
ошибок второго рода	7
Размер проверочной выборки	70

### Эксперимент

4

Обучающая выборка = 100, проверочная = 70. Количество правил = 5

Входные переменные Генеральный коэффициент ликвидности (собственный капитал/работающие активы), коэффициент мгновенной ликвидности (ликвидные обязательства+ защищенный капитал+ средства в резервном фонде суммарные обязательства, коэффициент фондовой капитализации прибыли (собственный капитал/ уставной фонд). Соответствующие результаты приводятся в таблице 4.

### Эксперимент 5

Обучающая выборка=100, проверочная =70. Количество правил = 5

Входные переменные Коэффициент ликвидности (собственный капитал/ работающие активы), коэффициент мгновенной ликвидности (ликвидные активы/ обязательства по требованию),

Кросс-коэффициент (суммарные обязательства/работающие активы)

Генеральный коэффициент ликвидности (ликвидные обязательства+ защищенный капитал+ средства в резервном фонде суммарные обязательства,

коэффициент фондовой капитализации прибыли (собственный капитал/ уставной фонд).

Коэффициент защищенности капитала.

Заметим, что эти показатели используются в методике Кромонава[3]. Соответствующие результаты приводятся в таблице 5.

**Таблица 5**

Результаты	TSK
Общее количество ошибок	9
%% ОШИБОК	13%
ошибок первого рода	1
ошибок второго рода	8
Размер проверочной выборки	70

В последующих экспериментах проводились сравнительные исследования эффективности прогнозирования с использованием классических четких методов и ННС. Соответствующие результаты приводятся в таблице 6.

**Таблица 6**

Методы	Общее количество ошибок	Ошибок первого рода	Ошибок второго рода	% % ошибок
ANFIS	7	1	6	10
TSK	5	0	5	7
Методика Кромонава	10	5	5	15
Методика АББ	10	2	8	15

Методика АББ – методика многоуровневого агрегированного показателя состояния банка Ассоциации белорусских банков.

Как следует из приведенных результатов, нечеткие методы на основе ННС TSK и ANFIS дают лучшие результаты в сравнении с общепринятыми четкими методами и методиками прогнозирования банкротства

## Выводы

В целом на основании проведенных многочисленных экспериментов были сделаны следующие выводы:

- 1) Из предложенных и рассмотренных методик нейронная сеть ННС TSK показывает наилучшие результаты в сравнении с ANFIS, методом Кромонава и методикой Ассоциации белорусских банков (АББ). Процент ошибок для TSK равен 7% , для ANFIS - 10%, четкие методы показали 15%.
- 2) Наиболее оптимальными входными показателями для прогнозирования с использованием ННС являются следующие: генеральный коэффициент ликвидности, коэффициент мгновенной ликвидности, кросс – коэффициент, коэффициент фондовой капитализации прибыли.
- 3) Период входных данных влияет на качество прогноза, а именно, чем ближе этот период к текущей дате, тем точнее прогноз.

---

### Благодарности

---

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

---

### Литература

---

1. Рейтингова оцінка комерційного банку. [http://www.nbu.gov.ua/Soc\\_Gum/Ekpr/2009\\_25/mescheryakov.htm](http://www.nbu.gov.ua/Soc_Gum/Ekpr/2009_25/mescheryakov.htm)
  2. Построение рейтинга банков с использованием методики расчета многоуровневого агрегированного показателя состояния банка <http://www.credit-rating.ua/ru/analytics/analytical-articles/>
  3. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах.- К.: Издательский дом «Слово», 2008.- 344с.
- 

### Информация об авторах

---

**Зайченко Юрий Петрович**, профессор, д.т.н., декан факультета «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14.

тел: 38044-406-83-93, e-mail: [ZaychenkoYuri@ukr.net](mailto:ZaychenkoYuri@ukr.net), [baskervil@voliacable.com](mailto:baskervil@voliacable.com)

**Войтенко Ольга**, бакалавр компьютерных наук, «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14.

## ПОСТРОЕНИЕ И ИССЛЕДОВАНИЯ НЕЧЕТКОЙ МОДЕЛИ ПОТРЕБЛЕНИЯ

Олег Пошелюжный

**Abstract:** В данной работе рассматривается неоклассическая задача теории потребления. Применяется нечеткий подход для построения и оптимизации функции полезности. Приводится иллюстративный пример.

**Keywords:** микроэкономика, теория потребления, функция полезности, методы оптимизации, нечеткие множества, принцип расширения.

**ACM Classification Keywords:** Decision Support & Artificial Intelligence

### Введение

В теории принятия решений часто возникают ситуации, в которых исходные условия задачи нечетко определены. Такие ситуации отображают недостаточную информированность лица, принимающего решение (ЛПР). Таким образом, для дальнейшего применения математических методов анализа и исследование все более сложных систем появилась потребность в использовании теории нечетких множеств. Одним из наиболее актуальных направлений этой теории является проблема принятия решений при нечетких условиях и критериях, которая привела к появлению нового направления в математическом программировании – нечеткого математического программирования (НМП).

### Классическая постановка

Пусть существует конечное число  $n, n \geq 1$ , товаров, которые имеют свойство количественной измеримости. Выбор потребителя характеризуется набором товаров  $x = (x_1, \dots, x_n) = (x_i)_{i=1, n}$ ,  $x_i \geq 0$ , где  $x_i$  — количество  $i$ -го товара, приобретенного потребителем. Тогда все возможные наборы товаров является точками пространства товаров  $X$  (множество альтернатив)  $X \subseteq R_+^n$ .

Выбор потребителем определенного набора товаров зависит не только от его потребностей, а частично и от его вкусов. Он характеризуется субъективным отношением предпочтения на множестве товаров, которое обозначим  $\succeq$ . В математической теории потребления считается, что отношение предпочтения удовлетворяет ряду аксиом, которые интерпретируют естественные свойства выбора потребителя и обеспечивают возможность развития полноценной теории. Основную группу аксиом представляют такие:

A1. Отношение предпочтения является полным квази порядком в пространстве товаров  $X$ , т.е. является бинарным отношением, для которого выполняются такие свойства [Пономаренко, 2004]:

- I. Отношение  $\succeq$  рефлексивное, т.е.  $x \succeq x, \forall x \in X$ ;
- II. Отношение  $\succeq$  транзитивное, т.е. из  $x \succeq y$  и  $y \succeq z$  для  $\forall x, y, z \in X$  вытекает, что  $x \succeq z$ ;
- III. Отношение  $\succeq$  полное, т.е. для  $\forall x, y \in X$  или  $x \succeq y$ , или  $y \succeq x$ .

A2. Отношение предпочтения  $\succeq$  непрерывное.

Тогда пара  $(X, \succeq)$ , т.е. пространство товаров  $X$  с отношением предпочтения  $\succeq$  определенного потребителя, называется полем предпочтений этого потребителя.

**Определение 1.** Числовая функция  $U: X \rightarrow R_+^1$ , определенная на  $X$ , называется функцией полезности, которая представляет отношение предпочтения  $\succeq$ , если для  $\forall x, y \in X$  выполняется  $(U(x) \geq U(y)) \Leftrightarrow (x \succeq y)$ .



Функция полезности потребителя выражает количественную меру ценности разных потребительских наборов товара из пространства товаров. Одним из основных результатов теории потребления и теории полезности в задачах принятия решений является теорема Ж. Дебре, которая задает достаточные условия существования функции полезности [Волошин, 2006].

### Нечеткая постановка задачи потребления

Процесс принятия решений как выбор наилучшей альтернативы из множества альтернатив  $X$ , может происходить при разной степени информированности ЛПР. На практике могут быть такие ситуации, когда у ЛПР отсутствует четкое представление об отношении предпочтения между всеми или некоторыми альтернативами, а можно лишь оценить степень выполнения того или другого предпочтения между парами альтернатив в виде числового значения  $\mu$ ,  $\mu \in [0, 1]$ .

Аналогично “обычным” бинарным отношением можно рассматривать понятие нечеткого бинарного отношения на множестве  $X$ , как нечеткое подмножество декартового произведения  $X \times X$  с функцией принадлежности  $\mu_R(x, y)$ ,  $0 \leq \mu_R(x, y) \leq 1$ , тогда, с помощью ЛПР, можно ввести нечеткое отношение предпочтения [Орловский, 1981].

**Определение 2.** Нечетким отношением предпочтения на множестве альтернатив  $X$  будем называть любое заданное на этом множестве нечеткое рефлексивное отношение  $R$ , которое характеризуется функцией принадлежности  $\mu_R : X \times X \rightarrow [0, 1]$ .

На нечеткое отношение предпочтения в теории потребления накладываются условия  $A1, A2$ .

Если  $R$  – нечеткое отношение предпочтения на множестве  $X$ , то для произвольной пары альтернатив  $(x, y) \in R$  значения  $\mu_R(x, y)$  следует понимать как степень выполнения предпочтения  $x \succeq y$ , далее нечеткое отношение предпочтения будем обозначать как  $x \succeq^\mu y$ . Используя определение 1, предложим следующее определение функции полезности при нечетком отношении предпочтения.

**Определение 3.** Числовая функция  $U : X \rightarrow R_+^1$ , определенная на  $X$ , называется функцией полезности, которая представляет нечеткое отношение предпочтения  $\succeq^\mu$ ,  $0 \leq \mu \leq 1$ , если для  $\forall x, y \in X$  выполняется  $\left( U(x) \geq^\mu U(y) \right) \Leftrightarrow \left( x \succeq^\mu y \right)$ .

Для ЛПР набор товаров  $x$  предпочтительнее набора  $y$  со степенью  $\mu$ , тогда и только тогда, когда число  $U(x)$  больше за число  $U(y)$  со степенью  $\mu$ , т.е. применяется понятие “субъективной вероятности” – “в  $\mu$  случаях набор  $x$  есть предпочтительнее набора  $y$ ”.

Согласно теории экономического поведения потребителя считается, что потребители выбирают наилучший относительно сравнительных преимуществ набор товаров, который они могут себе позволить. Тогда пусть ЛПР был осуществлен рациональный выбор нечеткого подмножества недоминированных альтернатив  $X^p = \left\{ x \in X \mid \mu^p(x) = 1 - \sup_{y \in X} (\mu_R(y, x) - \mu_R(x, y)) \right\}$ , где  $\mu^p(x)$  – есть степень, с которой альтернатива  $x$  не доминируется ни одной из альтернатив множества  $X$ .

Например, для человека потребление набора товаров  $x$ , которые оцениваются 1500 ккал. является более полезным за набор товаров  $y$ , которые дают 500 ккал.  $\left( 1500 \geq^{6/7} 500 \right) \Leftrightarrow \left( x \succeq^{6/7} y \right)$ . Тогда бывают случаи (один раз в неделю), когда набор  $y$  предпочтительнее набора  $x$ , т.е.  $\left( 500 \geq^{1/7} 1500 \right) \Leftrightarrow \left( y \succeq^{1/7} x \right)$ .

На основе определения 3 возможны следующие методы построения нечеткой функции полезности.

1. Функция полезности является обычной четкой функцией, сравнение нечетким.
2. ЛПР строит функцию полезности на основе нечетких параметров.
3. Функция полезности является нечеткой функцией, сравнение четким.
4. Функция полезности и сравнение являются нечеткими.
5. ЛПР задает нечеткую функцию полезности.

При построении нечетких моделей используются различные типы функций принадлежности – треугольная, трапециевидальная, колоколообразная, гауссовская [Зайченко, 2008].

Рассмотрим основные классы функций полезности на нечетком подмножестве недоминированных альтернатив  $X^p$ .

- I. Класс линейных функций полезности.

$\tilde{U}(x) = \sum_{i=1}^n a_i x_i$ ;  $x = (x_i)_{i=\overline{1,n}} \in R_+^n$ ,  $a_i \geq 0$ , где  $a_i$  – нечеткий параметр, определяется как множество аргументов, для которых функция принадлежности  $\mu_i(a_i) \geq \lambda$ ,  $i = \overline{1,n}$ , где  $\lambda = \mu_{\geq}(x, y)$ ,  $\forall y \in X$ .

- II. Класс мультипликативных функций полезности.

$\tilde{U}(x) = \prod_{i=1}^n x_i^{\alpha_i}$ ;  $x = (x_i)_{i=\overline{1,n}} \in R_+^n$ ,  $\alpha_i \geq 0$ , где  $\alpha_i$  – нечеткий параметр.

При применении полей предпочтений потребителей  $(X, \succeq)$  чаще всего наблюдается ситуация, когда выбор потребителя ограничен бюджетом. К бюджетным факторам принадлежат цена  $p_i$  на товар вида  $i$ ,  $i = \overline{1,n}$ , и уровень потребительского дохода  $I$ . Тогда потребитель при выборе набора товаров  $x$  должен учесть бюджетное ограничение  $xp = p_1 x_1 + \dots + p_n x_n \leq I$ .

Таким образом, допустимые наборы товаров в пространстве  $R_+^n$  образуют потребительский симплекс  $S_n = \{x \in R_+^n \mid xp \leq I\}$ , который является замкнутым выпуклым множеством в  $R_+^n$ .

При заданной нечеткой функции полезности и бюджетном ограничении задача потребления принимает следующий вид:

$$\tilde{U}(x) \rightarrow \max, \quad xp \leq I, \quad x \in R_+^n. \quad (1)$$

Задача (1) является задачей нечеткого математического программирования (НМП). Поскольку допустимое множество для этой задачи является выпуклым и компактным, то она имеет единственное решение  $\xi^* = x^*(p, I)$ , которое называется функцией спроса потребителя, с функцией принадлежности множеству альтернатив со степенью  $\mu_D(\xi^*) = \min\{\mu_G(\xi^*), \mu_C(\xi^*)\}$ , где  $\mu_G(\xi^*)$  – функция принадлежности нечеткой цели,  $\mu_C(x^*)$  – функция принадлежности ограничению.

На практике, в условиях динамического развития рынка и конкурентной среды, цены на товары и услуги, а также доходы потребителей не могут быть определены “точно”. Как правило, они колеблются в определенном диапазоне в краткосрочном периоде. Поэтому, в качестве их описания предлагается использовать “нечетко” заданные цены и “нечеткий” доход, с заданием приоритета (функции принадлежности). Тогда задачу (1) можно записать следующим образом:

$$\tilde{U}(x) \rightarrow \max, \quad x\tilde{p} \leq \tilde{I}, \quad x \in R_+^n, \quad (2)$$

где  $\tilde{U}(x)$  – нечеткая функция полезности потребителя,  $\tilde{p}$  – нечеткий вектор цен на товары,  $\tilde{I}$  – нечетких доход (капитал).

### Исследование нечеткой модели потребления

Определим функцию Лагранжа  $L$  для задачи (2):  $L(x, \lambda) = \tilde{U}(x) + \gamma(\tilde{I} - x\tilde{p})$ . Условие оптимального потребления принимает вид системы уравнений:

$$\begin{cases} \frac{\partial \tilde{U}}{\partial x_i}(x_1^*, \dots, x_n^*) - \gamma^* \tilde{p}_i = 0, i = \overline{1, n}, \\ \tilde{I} - \sum_{i=1}^n \tilde{p}_i x_i^* = 0, x_i > 0, i = \overline{1, n}. \end{cases} \quad (3)$$

Задача (2) является ЗНМП, когда нечетко описанные как параметры в ограничении задачи, так и в самой максимизирующей функции. Предложим метод решения данной задачи. Введем несколько определений.

**Определение 4.** Нечетким числом называется нечеткое множество  $A$ , определенное на множестве действительных чисел  $A \subseteq R$  с функцией принадлежности  $\mu_A : R \rightarrow [0, 1]$  и удовлетворяющее условиям [Орловский, 1981]:

1.  $\sup_{x \in R} \mu_A(x) = 1$ , т.е. нечеткое множество  $A$  является нормальным.
2.  $\mu_A[\lambda x_1 + (1 - \lambda)x_2] \geq \min\{\mu_A(x_1), \mu_A(x_2)\}$ ,  $\forall x_1, x_2 \in R, \forall \lambda \in [0, 1]$ , т.е. нечеткое множество  $A$  выпуклое.
3. Функция  $\mu_A(x)$  является непрерывной.

Пусть  $f : X \rightarrow Y$  - заданное отображение и пусть  $A$  - некоторое нечеткое подмножество множества  $X$ . Соответственно обобщению Лотфи А. Заде, образ  $A$  при отображении  $\varphi$  определяется как нечеткое подмножество множества  $Y$ , которое представляет собой совокупность пар вида  $(y, \mu_B(y))$ , где  $\mu_B : Y \rightarrow [0, 1]$  функция принадлежности образа. Тогда можно записать:

$$\mu_B(y) = \sup_{x \in f^{-1}(y)} \mu_A(x), y \in Y \quad (4)$$

где  $f^{-1}(y) = \{x \mid x \in X, f(x) = y\}$ ,  $\forall y \in Y$ .

Выражение (3) есть одним из определений принципа расширения Лотфи А. Заде, для расширения области применения нечетких множеств на отображение. Применим принцип расширения для операций над нечеткими числами.

Если существует некоторое четкое отображение  $f : X_1 \times X_2 \times \dots \times X_n \rightarrow Y$ , а также некоторые нечеткие множества  $A_1 \subseteq X_1, A_2 \subseteq X_2, \dots, A_n \subseteq X_n$ , то соответственно принципу расширения нечеткое число  $B = f(A_1, A_2, \dots, A_n)$  определяется функцией принадлежности:

$$\mu_B(y) = \begin{cases} \sup_{(x_1, \dots, x_n) \in f^{-1}(y)} \min(\mu_{A_1}(x_1), \dots, \mu_{A_n}(x_n)), & \text{if } f^{-1}(y) \neq \emptyset, \\ 0, & \text{if } f^{-1}(y) = \emptyset. \end{cases} \quad (5)$$

Это разрешает найти функцию принадлежности нечеткого числа, которое отвечает значению четкой функции от нечетких аргументов.

Рассмотрим иллюстративный пример. Пусть потребитель делает выбор между двумя наборами товаров, которые содержат два товара хлеб и молоко  $x = (6, 1)$  и  $y = (3, 2)$ , и считает, что  $x \stackrel{0,7}{\succeq} y$ . Функция полезности имеет вид  $U(x_1, x_2) = 2x_1^\alpha x_2^\beta$ . Найти функцию спроса, если цены товаров  $p_1 = 1, p_2 = 3$  и доход потребителя  $I = 12$ .  $\alpha, \beta$  - нечеткие числа с функциями принадлежности треугольного и трапециевидального вида:

$$\mu_A(\alpha) = \begin{cases} 0, & \alpha \leq 1/4; \\ 4\alpha - 1, & 1/4 \leq \alpha \leq 1/2; \\ 3 - 4\alpha, & 1/2 \leq \alpha \leq 3/4; \\ 0, & \alpha \geq 3/4. \end{cases} \quad \mu_B(\beta) = \begin{cases} 0, & \beta \leq 1/8; \\ 8\beta - 1, & 1/8 \leq \beta \leq 1/4; \\ 1, & 1/4 \leq \beta \leq 1/3; \\ 3 - 6\beta, & 1/3 \leq \beta \leq 1/2; \\ 0, & \beta \geq 1/2. \end{cases}$$

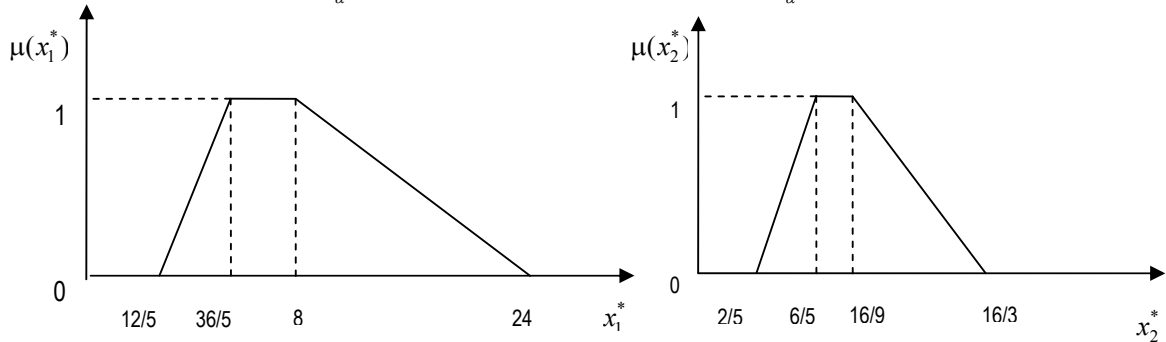
Задача оптимального поведения потребителя:

$$\begin{cases} U(x) = 2x_1^\alpha x_2^\beta \rightarrow \max, \\ p_1 x_1 + p_2 x_2 = I, \\ x_1, x_2 > 0. \end{cases}$$

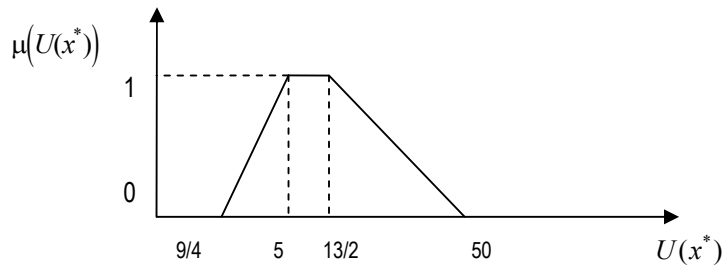
Применим условие оптимального потребления (3), тогда функция спроса первого товара  $x_1^* = \frac{\alpha I}{(\alpha + \beta)p_1}$ ,

функция спроса второго товара  $x_2^* = \frac{\beta I}{(\alpha + \beta)p_2}$ . Поскольку  $\alpha, \beta$  являются нечеткими числами, то

решение также является нечетким с функцией принадлежности  $\mu_D(x^*)$ . Найдем его с помощью принципа расширения (5):  $\mu_{A+B}(\eta) = \sup_{\alpha} \min\{\mu_A(\alpha), \mu_B(\eta - \alpha)\}$ ,  $\mu_{A/B}(\eta) = \sup_{\alpha} \min\{\mu_A(\eta\alpha), \mu_B(\alpha)\}$ , тогда



Построим функцию принадлежности для функции полезности. Таким образом,  $\mu(U(x)) = 0,9$ ,  $\mu(U(y)) = 0,75$ .



ЛПР анализирует величину интервала оптимальных значений полезности от определенного набора товаров  $\xi^* = x^*(p, I)$  в зависимости от степени достоверности интервала  $\mu(U)$  и принимает решение какое количество товаров ему приобрести.

---

### **Заключение**

Модели и методы нечеткого анализа ("Fuzzy Sets") дают возможность более адекватно описать задачу потребления в теории микроэкономики. Были построены нечеткие модели потребления (с нечеткими функциями полезности и нечеткими отношениями предпочтения), предложены алгоритмы получения нечеткого решения с использованием различных способов дефазификации решения. Рассмотрен иллюстративный пример с трапецеидальной функцией принадлежности.

---

### **Благодарности**

Автор благодарен проф. Волошину А.Ф. за постановку задачи и консультирование при написании статьи. Статья частично финансирована из проекта ITNEA XXI Института Информационных теорий и Приложений FOI ITNEA и Консорциума FOI Bulgaria ([www.itea.org](http://www.itea.org), [www.foibg.com](http://www.foibg.com)).

---

### **Литература**

- [Волошин, 2006] Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень.– К.: ВПЦ «Київський ун-т», 2006.– 304 с.
- [Пономаренко, 2004] Пономаренко О.І. та ін. Мікроекономіка. – Київ: Вища школа, 2004. – 262с.
- [Зайченко, 2008] Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. – Киев: Киевский политехнический институт, 2008. –344с.
- [Орловский, 1981] Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – Москва: Наука, 1981. – 206с.

---

### **Информация об авторе**

*Пошелюжный Олег* – Аспирант, Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: [oleg\\_posh@mail.ru](mailto:oleg_posh@mail.ru)

**Сфера научных интересов:** *принятие решений, системы поддержки принятия решений, математическая экономика, нечеткие модели и методы*

## ПАРАЛЛЕЛИЗАЦИЯ КООПЕРАТИВНЫХ МОДЕЛЕ-ОРИЕНТИРОВАННЫХ МЕТАЭВРИСТИК

Сергей Сиренко

**Аннотация:** Рассматривается кооперативный метаэвристический подход к решению задач комбинаторной оптимизации на основе модели-ориентированных алгоритмов. Его особенностью является решение задачи путем поиска (оптимизации) в пространстве моделей, который проводится на основе частных моделей, сформированных базовыми (составными) алгоритмами. Предлагается схема распараллеливания кооперативного метода, использующая присущий ему естественный параллелизм. Проведено исследование эффективности предлагаемой схемы параллелизации на основе анализов результатов вычислительного эксперимента.

**Ключевые слова:** комбинаторная оптимизация, модели-ориентированные методы, кооперативные метаэвристики, параллельные алгоритмы, оптимизация муравьиными колониями.

**ACM Classification Keywords:** G.1.6 [NUMERICAL ANALYSIS] Optimization, I.2.8 [ARTIFICIAL INTELLIGENCE]: Problem Solving, Control Methods, and Search – Heuristic methods, General Terms: Algorithms.

---

### Введение

В работе рассматривается метод построения кооперативных метаэвристик на основе модели-ориентированных алгоритмов [Гуляницкий, 2009]. Особенностью этого подхода к построению кооперативных методов решения задач комбинаторной оптимизации (ЗКО) является наличие управляющего алгоритма агрегирования (оптимизации) частных моделей, сформированных базовыми (составными) алгоритмами. В следующем разделе описана общая схема таких методов, далее излагаются детали разработанного подхода к реализации таких алгоритмов на многопроцессорных вычислительных системах. Результаты и обсуждение проведенного исследования эффективности предлагаемого способа распараллеливания на основе вычислительного эксперимента приведены в предпоследнем разделе.

---

### Построение кооперативных гибридных метаэвристик

Рассмотрим проблему построения гибридных кооперативных метаэвристик, которые базируются не на одном, а на нескольких модели-ориентированных алгоритмах; пусть  $K$  – число таких алгоритмов, которые назовем базовыми. Будем считать, что каждый такой алгоритм оперирует со своей моделью  $M_k, k = 1, \dots, K$ , а в результате их деятельности формируется один или несколько вариантов решения.

Рассмотрим одну из возможных реализаций. Предположим, что работа алгоритмов может осуществляться асинхронно, а их взаимодействие на итерации  $h$  происходит путем формирования вышестоящей процедурой агрегированной модели (метамодели)  $M^{h+1}$  как с учетом отдельных моделей  $M_1^h, \dots, M_K^h$  и сформированной на предыдущей итерации модели  $M^h$ , так и, возможно, сгенерированных алгоритмами решений. Таким образом, кооперативная метаэвристика оперирует несколькими моделями и осуществляет оптимизацию в пространстве моделей, с целью предоставления возможности базовым алгоритмам (или некоторым из них) генерировать наилучшие решения ЗКО.

Предлагается методология создания метаэвристических алгоритмов решения ЗКО, общая схема которых состоит из этапов, представленных на Рис. 1 [Гуляницкий и Сиренко; 2009]. Здесь для простоты изложения при формировании агрегированной модели (метамодели) не отражен явно фактор учета решений, сгенерированных базовыми алгоритмами.

Опишем ключевые этапы кооперативного алгоритма. После этапа инициализации осуществляется запуск всех базовых алгоритмов, каждый из которых итерационно и независимо генерирует решения и обновляет свою собственную модель. При выполнении условий обмена, текущие модели базовых алгоритмов (и, возможно, найденные ими варианты решений) используются для формирования новой агрегированной модели (метамоделей). При этом может также использоваться агрегированная модель с предыдущей итерации обмена, а сам процесс формирования может быть представлен в виде оптимизационной проблемы поиска наилучшего элемента в пространстве моделей.

```

procedure БазовыйАлгоритм( $k, M$ )
  Инициализация;
  while (не завершена работа) do
    ГенерированиеВариантовРешения ( $M_k, M$ );
    ОбновлениеМодели ( $M_k$ );
  end while
  return НаилучшееНайденноеРешение;
end procedure
procedure УправляющаяМетаэвристика()
  Инициализация;
  for  $k := 1$  to  $K$  do
    Запустить(БазовыйАлгоритм( $k, M^1$ ));
  end for
   $h := 1$ ;
  while (не выполняется условие завершения) do
    if (выполняется условие обмена) then
      for  $k := 1$  to  $K$  do
         $M_k^h :=$  ПолучитьМодель( $M_k$ );
      end for
       $M^{h+1} :=$  СформироватьМодель ( $M^h, M_1^h, \dots, M_K^h$ );
      for  $k := 1$  to  $K$  do
        РазослатьМодель( $k, M^{h+1}$ );
      end for
       $h := h + 1$ ;
    end if
  end while
  ЗавершитьРаботуАлгоритмов();
  return НаилучшиеНайденныеРешения;
end procedure

```

Рис. 1 Схема кооперативной модели-ориентированной метаэвристики

После этого агрегированная модель рассылается базовым алгоритмам, где может, как использоваться в сочетании с их собственной моделью, так и заместить ее полностью. При выполнении условий завершения работы кооперативная метаэвристика возвращает один или несколько наилучших из найденных базовыми алгоритмами вариантов решения.

В качестве базовых алгоритмов могут выступать отдельные экземпляры одного алгоритма, в этом случае схема называется гомогенной. Экземпляры алгоритмов могут иметь различные значения параметров. В случае гетерогенной схемы (когда базовые алгоритмы не основываются на одном алгоритме/методе), модели базовых алгоритмов должны быть согласованы, т.е. должен существовать способ преобразования значений параметров и/или структур моделей между собой.

Управляющая метаэвристика может формировать на одной итерации не одну, а сразу несколько разных агрегированных моделей (метамоделей). Это целесообразно как в случае гетерогенной схемы, когда

базовые алгоритмы могут использовать модели разной структуры, так и когда требуется исследовать несколько различных моделей одним алгоритмом.

При использовании данной методологии необходимо определить:

1. тип(ы) базовых модели-ориентированных алгоритмов;
2. способ использования базовыми алгоритмами агрегированной модели (метамодели);
3. условия обмена (схема коммуникации);
4. способ формирования агрегированной модели (метамодели);
5. условия завершения работы кооперативной метаэвристики.

Предполагается, что применение разработанной методологии позволит диверсифицировать работу базовых алгоритмов, что уменьшает вероятность завершения процесса поиска в областях, не содержащих глобально-оптимальное решение. Обмен информацией между базовыми алгоритмами позволяет надеется на повышение эффективности процесса поиска за счет проявления синергетических эффектов. Таким образом, выбор конкретной схемы обмена информацией (способа кооперирования) определяет баланс между интенсификацией и диверсификацией поиска.

Разработанная схема может также служить основой для классификации гибридных алгоритмов КО, например, на основе параметризации таких показателей, как синхронность выполнения и схемы взаимодействия базовых алгоритмов, их однотипность, количество используемых моделей и методы формирования агрегированных моделей, степень использования предыстории.

В этой работе исследуется применение в качестве процедуры агрегирования схемы взвешенного суммирования значений параметров (феромонных значений) моделей базовых алгоритмов  $\tau_1, \dots, \tau_K$  (для простоты изложения будем считать, что значения параметров нормированы). Веса вычисляются на основе значений целевой функции на текущих наилучших вариантах решений  $f_1^{opt}, \dots, f_K^{opt}$  и средних значений целевой функции  $f_1^{avg}, \dots, f_K^{avg}$  по найденным базовыми алгоритмами решениям на последней завершенной итерации по следующему соотношению:

$$\omega_k = \frac{0.2}{K} + 0.6 \frac{f_{\max}^{opt} - f_k^{opt}}{\sum_{i=1}^K (f_{\max}^{opt} - f_i^{opt})} + 0.2 \frac{f_{\max}^{avg} - f_k^{avg}}{\sum_{i=1}^K (f_{\max}^{avg} - f_i^{avg})}, k = 1, \dots, K; \quad (1)$$

$$f_{\max}^{opt} = \max_{k=1, \dots, K} f_k^{opt}, f_{\max}^{avg} = \max_{k=1, \dots, K} f_k^{avg}.$$

Поскольку исследовалось влияние наличия обмена информацией в кооперативной схеме на эффективность, то в качестве условия завершения было выбрано выполнение базовыми алгоритмами определенного количества итераций  $IT_{\max}$ .

### Распараллеливание кооперативной метаэвристической схемы

Эффективное распараллеливание метаэвристик является нетривиальной задачей и стало отдельной областью исследования [Parallel Metaheuristics, 2006]. Представленная кооперативная схема обладает естественным параллелизмом – она включает набор подчиненных алгоритмов выполняющихся одновременно со сравнительно низким обменом информации между ними. В общем случае наиболее целесообразным представляется не разбивать базовые алгоритмы между процессорами, чтобы минимизировать накладные расходы коммуникации. Поэтому мы предлагаем синхронную схему «главный-подчиненный» для распараллеливания кооперативной схемы, использующий представило агрегирования (1) (Рис. 2). Каждому процессору присваивается один или несколько базовых алгоритмов. Один процессор выделяется как главный и дополнительно выполняет операций управляющей метаэвристики.

Предлагается согласованная схема распределения  $K$  базовых алгоритмов на  $P$  процессорах ( $K \geq P$ ), который использует выполнение базовыми алгоритмами заданного количества  $IT$  итераций, как условие



выполнения обмена и агрегирования. В предлагаемой схеме все базовые модели берутся для агрегации с одной и той же итерации базовых алгоритмов.

Процедура агрегирования также частично распараллелена. Подчиненные процессоры отправляют главному информацию об их текущих лучших и средних значениях целевой функции. Главный процессор вычисляет веса и отправляет их подчиненным. Они нормализуют и суммируют свои базовые модели (если им присвоено больше одного основного алгоритма) и отправляют результат главному процессору. Он завершает процесс агрегирования и передает агрегированную модель подчиненным процессорам.

### Вычислительный эксперимент

Поскольку теоретическое исследование метаэвристических алгоритмов решения ЗКО крайне редко позволяет получать практически применимые результаты, принято анализировать показатели эффективности путем проведения вычислительных экспериментов. С этой целью обычно используют "классические" модели комбинаторной оптимизации – такие, например, как задача коммивояжера (ЗК) [Hoos and Stützle, 2005], которая состоит в поиске минимального гамильтонового цикла в полном взвешенном графе.

Проведенный эксперимент состоял в исследовании эффективности разработанной схемы распараллеливания для кооперативного алгоритма на базе алгоритмов оптимизации муравьиными колониями [Гуляницкий и Сиренко; 2009]. В качестве базовых алгоритмов при реализации были использованы экземпляры известного алгоритма Max-Min Ant System (MMAS) с реинициализацией феромонных значений [Stützle and Hoos, 2000], реализация которых взята из пакета ACOTSP [Stützle, 2004]. Феромонные значения реинициализируются, если решения сгенерированные на одной итерации достаточно близки друг к другу и на протяжении заданного количества итераций алгоритмом не было найдено улучшения. Модель в этих алгоритмах ОМК для ЗК представлена в виде квадратной матрицы, элементами которой являются феромонные значения. В алгоритме MMAS используются динамические ограничения  $\tau_{\max}, \tau_{\min}$  на значения элементов феромонной матрицы (параметров модели), поэтому перед агрегированием матриц осуществлялось масштабирование элементов с приведением их значений к отрезку  $[0, 1]$ :  $\tau_i^{norm} = (\tau_i - \tau_{\min}) / (\tau_{\max} - \tau_{\min})$ .

```

procedure ПодчиненныйПроцесс(pr_id)
     $K_{pr\_id} = K \text{ div } P$ ;
    if ( $P \bmod K \geq pr\_id$ ) then
         $K_{pr\_id} = K_{pr\_id} + 1$ ;
    end if
    Инициализация( $M, M_1, \dots, M_{pr\_id}$ );
    while (не завершена работа) do
        for  $i := 1$  to  $IT$  do
            for  $k := 1$  to  $K_{pr\_id}$  do
                ГенерированиеВариантовРешения( $M_k, M$ ); ОбновлениеМодели( $M_k$ );
            end for
        end for
        ОтправитьГлавномуПроцессу( $f_1^{opt}, f_1^{avg}, \dots, f_{K_{pr\_id}}^{opt}, f_{K_{pr\_id}}^{avg}$ );
        ПолучитьВеса( $\omega_1, \dots, \omega_{K_{pr\_id}}$ );
         $M^{temp} :=$  НормироватьИАгрегировать( $M_1, \omega_1, \dots, M_{K_{pr\_id}}, \omega_{K_{pr\_id}}$ );
        ОтправитьГлавномуПроцессу( $M^{temp}$ );
        ПолучитьМодель( $M$ );
    end while

```

```

end procedure
procedure ГлавныйПроцесс()
     $K_1 = K \text{ div } P$ ;
    Инициализация ( $M, M_1, \dots, M_{pr\_id}$ );
     $h := 1$ ;
    while (не выполняется условие завершения) do
        for  $i := 1$  to  $IT$  do
            for  $k := 1$  to  $K_1$  do
                ГенерированиеВариантовРешения ( $M_k, M$ ); ОбновлениеМодели ( $M_k$ );
            end for
        end for
        ПолучитьЗначения( $f_{K_1+1}^{opt}, f_{K_1+1}^{avg}, \dots, f_K^{opt}, f_K^{avg}$ );
        ВычислитьВеса( $\omega_1, \dots, \omega_K$ );
        РаспределитьМеждуПодчиненнымиПроцессами( $\omega_{K_1+1}, \dots, \omega_K$ );
         $M_1^{temp} :=$  НормироватьИАгрегировать( $M_1, \omega_1, \dots, M_{K_1}, \omega_{K_1}$ );
        ПолучитьЧастичныеМодели ( $M_2^{temp}, \dots, M_K^{temp}$ );
         $M^{h+1} :=$  Агрегировать( $M^h, M_1^{temp}, \dots, M_K^{temp}$ );
        РазовлатьМодель ( $M^{h+1}$ );
         $h := h + 1$ ;
    end while
    ЗавершитьРаботуАлгоритмов();
end procedure

```

Рис. 2 Схема параллельной кооперативной модели-ориентированной метаэвристики

Параметры базовых алгоритмов устанавливались равными стандартными значениям, рекомендуемыми в литературе для ЗК [Dorigo and Stützle, 2004], т.е. дополнительная их оптимизация для конкретного набора тестовых ЗК не проводилась. Количество муравьев составляло 25, коэффициент испарения феромона  $\rho = 0.5$ , в псевдослучайном пропорциональном правиле выбора, которое применяли муравьи при построении решений, были такие значения параметров:  $\alpha = 1$ ,  $\beta = 2$ . В качестве деятельности демона в алгоритмах ОМК ко всем построенным муравьями решениям применялся алгоритм простого локального поиска 3-opt [Hoos and Stützle, 2005], реализация которого также взята из пакета ACOTSP [Stützle, 2004]. Формирование управляющей метаэвристикой агрегированной модели осуществлялось после каждых 60 итераций, выполненных базовыми алгоритмами.

Тесты проводились на кластере СКИТ-3 Института кибернетики им. В.М. Глушкова. Была использована часть кластера с узлами, состоящими из двух четырех-ядерных процессоров Intel Xeon 5345 EM64T (2,2 ГГц на ядро) с 16 Гб оперативной памяти на узел. Алгоритмы были запрограммированы на языке C++. Для организации обмена сообщениями в параллельной версии использовалась реализация стандарта MPI версии OpenMPI 1.2.4.

Исследовались кооперативные алгоритмы, состоящие из 2, 5, 10 и 25 копий алгоритма MMAS. Три ЗК att532, u1432, pr2392 размерностей 532, 1432 и 2392 городов соответственно были отобраны, и замерялось среднее время необходимое алгоритмам для того, что бы достигнуть точности 0,75% относительно известного оптимального решения. Средние значения вычислялись на основе 30 попыток.

В табл. 1 приведены эффективности алгоритмов на 2, 5, 10 и 25 процессорах (ядер). Тут  $K$  – число базовых алгоритмов, а  $P$  – число задействованных процессоров. Конфигурации, для которых  $K < P$  не рассматривались. Эффективности вычисляются на основе времени относительного ускорения (ускорение параллельного кода относительно того же кода запущенного на одном ядре).

Таблица 1 Результаты решения ЗК

Задача	att532				u1432				pr2392				
	<i>K</i>	2	5	10	25	2	5	10	25	2	5	10	25
<i>P</i>													
2		0.88	1.10	0.89	1.00	0.86	1.07	0.91	0.98	0.85	1.05	0.97	0.95
5		–	0.81	0.84	0.87	–	0.67	0.70	0.87	–	0.48	0.53	0.64
10		–	–	0.67	0.97	–	–	0.59	0.94	–	–	0.40	0.61
25		–	–	–	0.65	–	–	–	0.52	–	–	–	0.34

В целом ускорение параллельного алгоритма близко к линейному для меньших значений количества процессоров и базовых алгоритмов. Однако при возрастании размерности наблюдается тенденция к значительному снижению эффективности параллельного алгоритма. Также отметим, что более целесообразно присваивать одному процессору (ядру) несколько базовых алгоритмов для снижения расходов ресурсов на обмен сообщениями.

Для 2 ядер и 5 базовых алгоритмов наблюдается сверхлинейное ускорение. Это факт объясняется тем, что в этом случае главному ядру присваивается 2, а подчиненному – 3 базовых алгоритма. И, поскольку накладные расходы в случае двух ядер невелики, это позволяет главному ядру находить решение нужного качества более чем в два раза быстрее по сравнению с последовательным алгоритмом.

Проведенный эксперимент показал, что кооперативная модели-ориентированная схема может быть распараллелена эффективно.

### Заключение

В работе исследуются возможности параллелизации кооперативной модели-ориентированной метаэвристической схемы. Рассмотрен подход, в котором каждому процессору присваивается один или более базовых алгоритмов. Экспериментальное исследование на задачах коммивояжера с известными решениями показало, что разработанный параллельный алгоритм позволяет эффективно использовать естественный параллелизм присущий кооперативной схеме.

### Благодарности

Статья частично профинансирована в рамках проекта **ITHEA XXI** Института Информационных Теории и Приложений FOI ITHEA и ADUIS Ukraine ([www.ithea.org](http://www.ithea.org), [www.aduis.com.ua](http://www.aduis.com.ua)).

### Список литературы

- [Hoos and Stützle, 2005] H.H. Hoos, T. Stützle. Stochastic Local Search: Foundations and Applications. San Francisco: Morgan Kaufmann Publ, 2005. 658 p.
- [Parallel Metaheuristics, 2006] Parallel Metaheuristics: A New Class of Algorithms (eds. E. Alba). (2006) Wiley. 576 p.
- [Stützle, 2004] T. Stützle. ACOTSP, Version 1.0. <http://www.aco-metaheuristic.org/aco-code>. 2004.
- [Гуляницький, 2009] Л.Ф. Гуляницький. Розробка кооперативних метаевристик. In: Abstract of Int. Conf. "Problems of Decision Making under Uncertainties (PDMU-2009)" (April 27-30, 2009, Skhidnytsia, Ukraine). Kyiv, 2009. pp. 90–91.
- [Гуляницький и Сиренко, 2009] Л.Ф. Гуляницький, С. Сиренко. Кооперативные модели-ориентированные метаэвристики для задач комбинаторной оптимизации. In: Intelligent Support of Decision Making (Eds. Krassimir Markov et al.). Sofia: ITHEA, 2009. pp. 165–172.

### Информация об авторе



**Сергей Сиренко (Sirenko)** – аспирант, Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: [s.sirenko@gmail.com](mailto:s.sirenko@gmail.com)

## ПРАКТИЧЕСКОЕ ИСПОЛЬЗОВАНИЕ ВИРТУАЛЬНОЙ ЛАБОРАТОРИИ ДЛЯ РАСЧЕТА НАДЕЖНОСТИ ПРИБОРОВ И РЕЗЕРВИРОВАННЫХ СИСТЕМ

Александр Палагин, Владимир Романов, Игорь Галелюка,  
Крассимир Марков, Виталий Величко, Александра Ковырёва,  
Петер Станчев, Крассимира Иванова, Илия Митов

**Аннотация:** В статье рассмотрено практическое применение программных модулей виртуальной лаборатории автоматизированного проектирования для расчета параметров надежности резервированных систем и портативного прибора "Флоратест", малая серия которого выпущена в Институте кибернетики им. В.М.Глушкова НАН Украины. Приведенные примеры расчетов показывают возможность использования виртуальной лаборатории для расчета резервированных систем разной сложности и приборов разнообразного назначения.

**Ключевые слова:** виртуальная лаборатория; автоматизированное проектирование; расчет надежности; резервирование надежности; распределенная система.

**ACM Classification Keywords:** J.6 Computer-Aided Engineering – Computer-Aided Design (CAD); K.4.3 Organizational Impacts – Computer-Supported Collaborative Work.

---

### Введение

Для портативных приборов широкого применения важным является, кроме параметров назначения, высокая надежность. Это, прежде всего, связано с условиями эксплуатации этих приборов. Чаще всего портативные приборы используются в полевых условиях с высокими перепадами температур, давления, влажности и т.п. На них влияют тряска и вибрации при транспортировке, ударная нагрузка при эксплуатации и т.п. Кроме того, в удаленных районах эксплуатации, как правило, отсутствуют сервисные центры обслуживания портативных приборов конкретного назначения. Поэтому показатели надежности (вместе с показателями назначения) являются наиболее важными эксплуатационными показателями портативного прибора или средства. Это обусловило создание в составе виртуальной лаборатории автоматизированного проектирования [Palagin, 2009], которая разработана совместно Институтом кибернетики им. В.М. Глушкова НАН Украины и Институтом математики и информатики БАН, модулей расчета показателей надежности. С целью повышения надежности на практике часто используют резервирование, которое заключается в параллельном подключении двух (дублированная система) или трех (утроенная система) элементов. Виртуальная лаборатория не имеет ограничений относительно анализа надежности в зависимости от количества уровней резервирования.

---

### Расчет надежности резервированных систем

В нашем случае мы будем рассматривать невозстанавливаемые системы. Отказом системы будем считать выход из строя (отказ) всех элементов. Имеется в виду, что резервирование в таких системах является "горячим". Использование "горячего" резервирования заведомо допускает появления ряда недостатков, которые проявляются в увеличении общей стоимости всей системы, повышении энергопотребления и резком снижении вероятности безотказной работы через некоторое время. Чем сложнее система, тем это время меньше. То есть, "горячее" резервирование гарантирует высокую вероятность безотказной работы системы в коротком промежутке времени.

В виртуальной лаборатории нами реализован расчет вероятности безотказной работы резервированных систем. Для этого были проанализированы методики классической теории надежности [Гнеденко, 1965, Стрельников, 2004], которые потом были модифицированы и реализованы в виде программного модуля.

Для получения аналитического выражения определения вероятности безотказной работы резервированных систем мы использовали аппарат классической теории вероятности, а именно теоремы

сложения и умножения вероятностей. В нашем случае необходимо определить вероятность события  $\Pi$ , которое состоит в том, что система, которая представляет собой параллельное соединение двух элементов проработает безотказно в интервале времени  $(0, t)$ . Согласно [Азарсков, 2004] система проработает безотказно в интервале времени  $(0, t)$ , если произойдет одно из несовместимых событий:

$\Pi_1$  – за время  $t$  не откажет не один из элементов;

$\Pi_2$  – за время  $t$  откажет элемент "2", а элемент "1" проработает без отказа;

$\Pi_3$  – за время  $t$  откажет элемент "1", а элемент "2" проработает без отказа.

На основе теоремы сложения вероятностей вероятность безотказной работы системы за время  $t$  будет равна

$$P(t) = P(\Pi_1) + P(\Pi_2) + P(\Pi_3) \quad (1)$$

где  $P(t)$  – вероятность безотказной работы резервированной системы;  $P(\Pi_1)$ ,  $P(\Pi_2)$ ,  $P(\Pi_3)$  – вероятности появления событий  $\Pi_1$ ,  $\Pi_2$ ,  $\Pi_3$  соответственно.

Согласно теореме умножения вероятностей нами определена вероятность каждого из трех событий:

$$\begin{aligned} P(\Pi_1) &= P_1(t) \cdot P_2(t) \\ P(\Pi_2) &= P_1(t) \cdot (1 - P_2(t)) \\ P(\Pi_3) &= (1 - P_1(t)) \cdot P_2(t) \end{aligned} \quad (2)$$

где  $P_1(t)$ ,  $P_2(t)$  – вероятность безотказной работы элементов "1" и "2" соответственно.

Подставив (2) в (1), получим выражение для определения вероятности безотказной работы дублированной системы

$$P(t) = P_1(t) + P_2(t) - P_1(t) \cdot P_2(t) \quad (3)$$

Поскольку мы полагаем, что элементы 1 и 2 имеют одинаковую надежность, то выражение (3) примет следующий вид

$$P(t) = 2P_e(t) - P_e(t)^2 \quad (4)$$

где  $P_e$  – вероятность безотказной работы одного из параллельных элементов.

Получение аналитического выражения для вычисления вероятности безотказной работы утроенной системы происходит аналогично описанной выше последовательности на основе теорем сложения и умножения вероятностей. Конечное выражение для определения вероятности безотказной работы утроенной системы будет иметь следующий вид:

$$P(t) = P_e(t)[3 - 3P_e(t) + P_e(t)^2] \quad (5)$$

Приведенные выше расчеты вероятностей безотказной работы резервированных систем адаптированы и реализованы нами в виртуальной лаборатории в виде программных модулей расчета параметров надежности.

Кроме непосредственных расчетов вероятностей безотказной работы резервированных систем адаптированные методики позволяют решать также другую задачу. Данная задача состоит в определении степени резервирования системы, что даст возможность достичь заданных пользователем показателей безотказности работы.

Нами совершен расчет вероятности безотказной работы для обычной и резервированной систем согласно адаптированным методикам для проверки их соответствия классической теории надежности. Под резервированными системами понимаем дублированную и утроенную системы, которые рассмотрены выше. Принято, что система состоит из одинаковых микроселектронных элементов, которые изготовлены,

например, по технологии "Bipolar < 2.5  $\mu\text{m}^2$ ". Расчеты осуществлены для 4 систем, которые состоят из 1 тыс., 10 тыс., 20 тыс. и 100 тыс. элементов соответственно. Доверительные границы равны 60 %.

Результаты расчетов вероятности безотказной работы для четырех типов систем с использованием экспоненциальной модели распределения вероятностей отказов приведены в табл. 1, а с использованием модели DN-распределения вероятностей отказов – в табл. 2.

Таблица 1. Вероятности безотказной работы систем, полученные с использованием экспоненциальной модели распределения вероятностей отказов

Количество элементов в системе, тыс.	Вероятность безотказной работы системы		
	обычная	дублированная	утроенная
1	0,924	0,994	0,999
10	0,672	0,893	0,964
20	0,204	0,367	0,496
100	0,00035	0,00071	0,00107

Таблица 2. Вероятности безотказной работы систем, полученные с использованием модели DN-распределения вероятностей отказов

Количество элементов в системе, тыс.	Вероятность безотказной работы системы		
	обычная	дублированная	утроенная
1	0,554	0,801	0,911
10	0,123	0,231	0,325
20	0,060	0,116	0,169
100	0,005	0,00997	0,01492

По результатам расчетов можно сделать вывод про соответствие полученных результатов данным, которые рассчитаны в ручном режиме с использованием зависимостей теории надежности [Азарков, 2004].

Для анализа вероятности безотказности работы резервированных систем с помощью программного модуля расчета параметров надежности нами были построены зависимости вероятности безотказной работы обычной, дублированной и утроенной систем, которые состоят из 1 тыс. и 100 тыс. элементов, от времени эксплуатации. Доверительные границы были равны 60 %. На рис. 1 и рис. 2 приведены зависимости, полученные с помощью методики на основе экспоненциального закона распределения вероятности отказов. На рис. 3 и рис. 4 приведены зависимости, полученные с помощью методики на основе DN-распределения вероятности отказов.

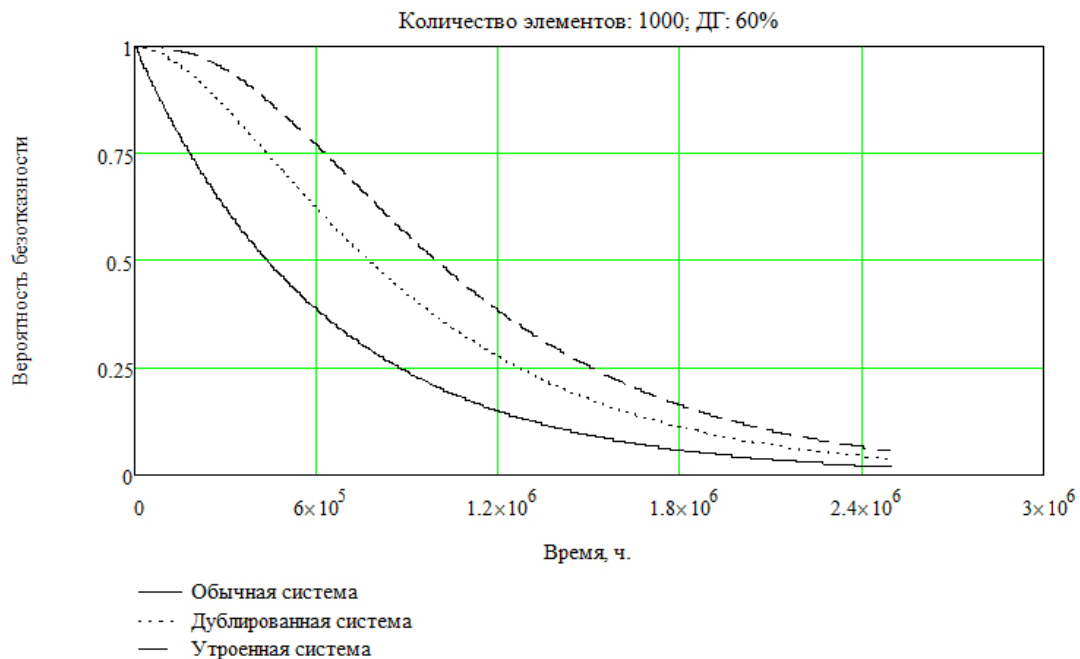


Рисунок 1. Зависимость вероятностей безотказной работы обычной и резервированной систем (1 тыс. элем.) от времени эксплуатации, экспоненциальный закон

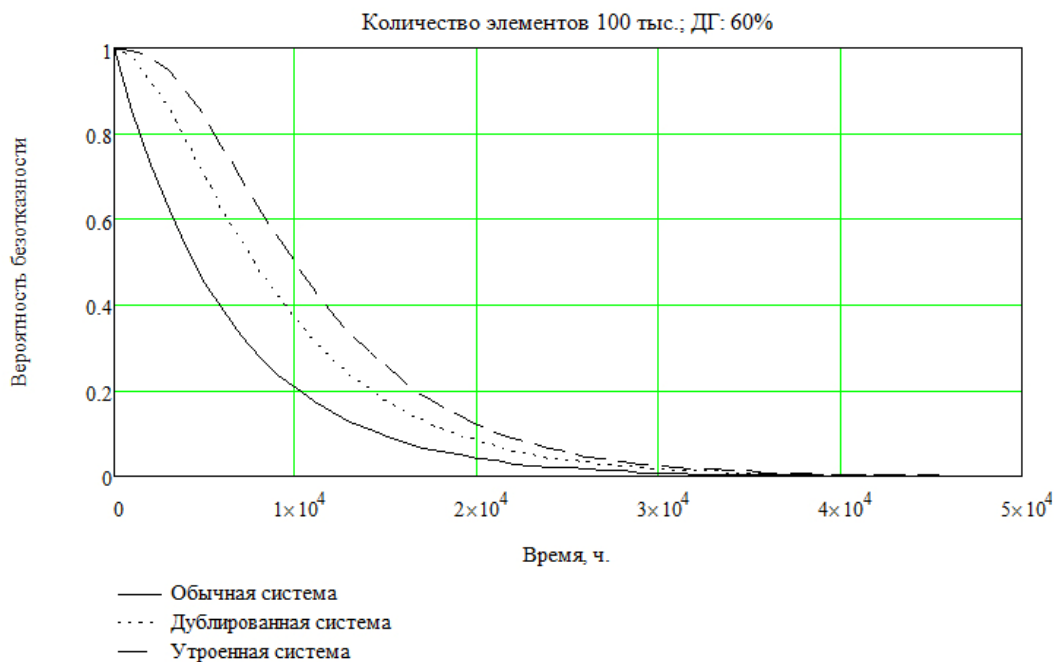


Рисунок 2. Зависимость вероятностей безотказности работы обычной и резервированных систем (100 тыс. элем.) от времени эксплуатации, экспоненциальный закон

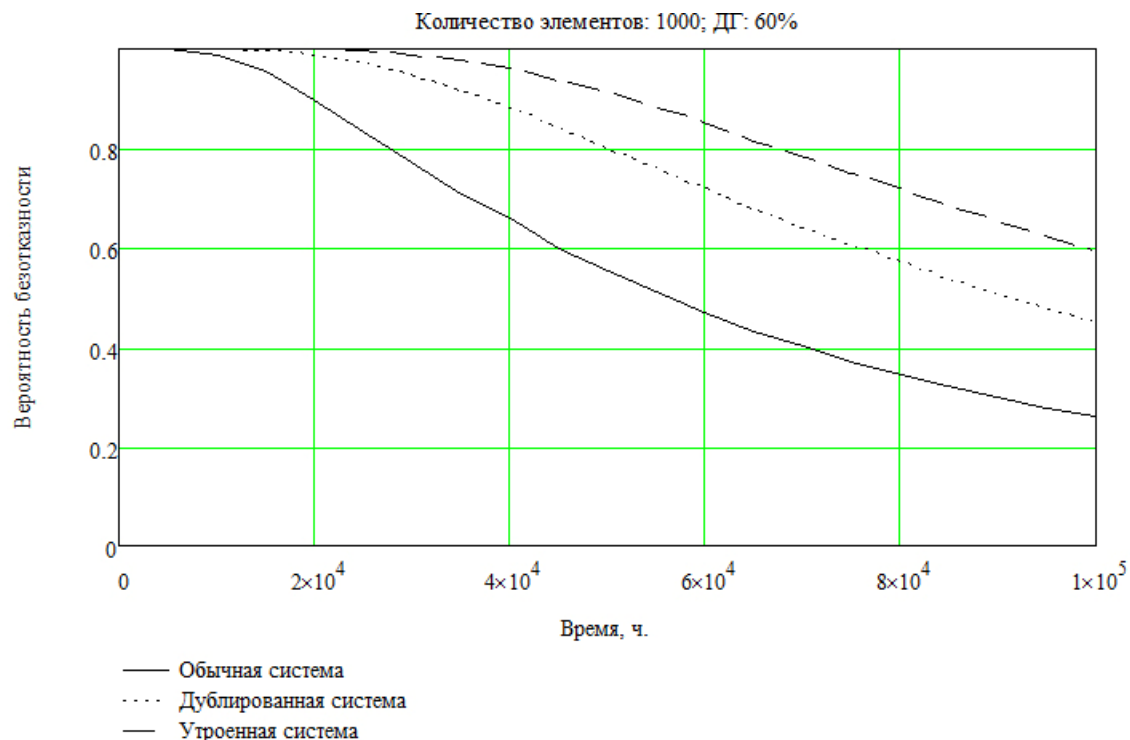


Рисунок 3. Зависимость вероятностей безотказности работы обычной и резервированных систем (1 тыс. элем.) от времени эксплуатации, DN-распределение

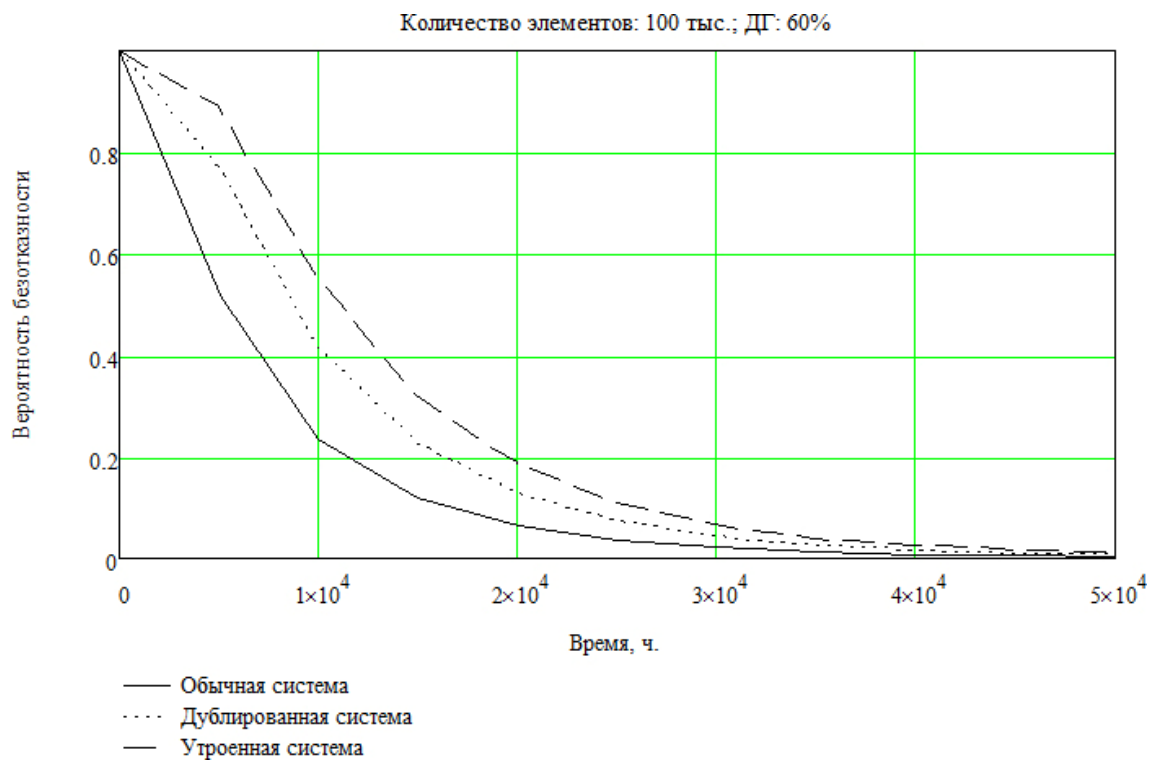


Рисунок 4. Зависимость вероятностей безотказности работы обычной и резервированных систем (100 тыс. элем.) от времени эксплуатации, DN-распределение



Проанализировав графики (рис. 1 и рис. 2), мы четко определили, что при расчете с помощью методики на основе экспоненциального закона распределения вероятность безотказной работы систем, которые состоят из 1 тыс. и 100 тыс. элементов, после начала эксплуатации начинает резко уменьшаться. Это согласуется с классической теорией надежности [Азарсков, 2004, Стрельников, 2002]. Резервирование систем помогает поддержать вероятность безотказной работы на уровне 0,999 в начале эксплуатации на протяжении определенного периода времени. Для системы из 1 тыс. элементов этот период для дублированной системы равен 10 тыс. часов (или 417 дней), а для утроенной системы – 20 тыс. часов (или 833 дня). Для системы из 100 тыс. элементов эти периоды составляют приблизительно 1000 часов (или 42 дня) и 2000 часов (или 83 дня) соответственно для удвоенной и утроенной систем.

Проанализировав графики (рис. 3 и рис. 4), мы четко видим, что при DN-распределении вероятность безотказной работы систем, которые состоят как в предыдущем случае из 1 тыс. та 100 тыс. элементов, после начала эксплуатации ведет себя различным образом. Для системы из 1 тыс. элементов даже обычная система после начала эксплуатации имеет небольшую область (приблизительно 7000 часов), на которой вероятность безотказной работы составляет 0,999. При резервировании такой системы периоды времени, на протяжении которых вероятность безотказной работы равна 0,999, значительно возрастают. Для дублированной системы он равен 20 тыс. часов (или 833 дня), а для утроенной системы – 27 тыс. часов (или 1125 дней). Вероятность безотказной работы системы из 100 тыс. элементов ведет себя для трёх систем почти одинаково, а именно, начинает резко уменьшаться уже от начала первого включения. Разница состоит лишь в крутизне кривой. То есть, период времени, в котором вероятность безотказной работы равна 0,999, является чрезвычайно малым. Для улучшения надежность в таком случае лучше увеличивать уровни резервирования системы.

---

### Расчет надежности проектируемого прибора

---

Виртуальная лаборатория была использована нами для расчета параметров надежности портативного прибора для экспресс-диагностики состояния растений "Флоратест" [Romanov, 2007], который был разработан в Институте кибернетики имени В.М.Глушкова НАН Украины. Исходными данными для расчетов послужили принципиальная электрическая схема базового блока прибора "Флоратест" и параметры надежности микросэлектронных элементов, которые взяты из официальных сайтов компаний-производителей этих элементов. Расчет параметров надежно выполнен с помощью двух рассмотренных методик: [Palagin, 2010]:

- 1) на основе экспоненциального закона распределения вероятностей отказов;
- 2) на основе DN-распределения вероятностей отказов.

Расчеты проведены нами для доверительных границ 60% и 90%.

Портативный прибор "Флоратест" состоит из двух основных блоков: базового блока и выносного сенсора. Согласно принципиальной электрической схеме прибора нами был создан перечень всех элементов базового блока прибора. Дальше все элементы базового блока были разделены нами на 2 группы:

- 1) параметры надежности элементов задаются в единицах FIT (отказ на  $10^9$  часов наработки);
- 2) надежность элементов определяется количеством циклов срабатываний (кнопки, переключатели и т.п.).

Параметры надежности отдельного элемента первой группы рассчитываются на основе сведенных данных по надежности технологии изготовления этого элемента. Параметры надежности элементов второй группы рассчитываются путем сравнения необходимого количества цикла срабатываний на протяжении гарантированного срока работы портативного прибора с гарантированным производителем элемента количеством циклов срабатываний.

Приведем алгоритм расчета параметров надежности базового блока портативного прибора "Флоратест" с помощью модуля расчетов параметров надежности, который входит в состав виртуальной лаборатории:

- 1) Задать количество типов микросэлектронных компонентов и общее количество микросэлектронных компонентов в приборе.
- 2) Выбрать из списка название каждого типа микросэлектронного компонента и ввести количество компонентов каждого типа.
- 3) Проверить правильность введенной информации.
- 4) Нажать кнопку "Рассчитать" для расчета параметров надежности базового блока прибора "Флоратест".
- 5) На экран монитора будут выведены рассчитанные двумя методиками параметры надежности каждого типа микросэлектронных компонентов и параметры надежности базового блока прибора.

Следует заметить, что по приведенному выше алгоритму совершается расчет параметров надежности базового блока прибора "Флоратест" на основе параметров надежности элементов первой группы.

Согласно введенным данным нами получены следующие результаты расчета параметров надежности базового блока портативного прибора "Флоратест" (см. табл. 3).

Таблица 3. Результаты расчета параметров надежности базового блока прибора "Флоратест"

Закон распределения вероятностей отказов	Доверительная граница			
	60 %		90 %	
	часов	лет	часов	лет
Экспоненциальный	1353102	154,5	1119396	127,8
DN	136711	15,6	133097	15,2

Как было сказано выше, параметры надежности элементов второй группы рассчитываются путем сравнения необходимого количества циклов срабатываний на протяжении гарантированного срока работы портативного прибора с гарантированным производителем элемента количеством циклов срабатываний.

Для кнопки, которая входит в состав элементов второй группы, производителем (компания "Schurter") предоставляется гарантия в среднем от 800 тыс. до 1,5 млн. циклов нажиманий в зависимости от тока. Для совершения одного измерения необходимо нажать на каждую из кнопок дважды. Для переноса данных каждые 20 измерений из прибора в ПК необходимо также дважды нажать на каждую из кнопок. Исходя из того, что с помощью прибора "Флоратест" планируется делать в среднем до 100 измерений в день, несложно рассчитать ежедневное количество нажатий одной кнопки. Следовательно, за один день пользователь нажмет каждую из кнопок 210 раз. Соответственно в год каждая кнопка будет нажата 76,5 тыс. раз и её гарантийный срок службы будет равен от 10 до 20 лет.

На основе проведенных расчетов можно утверждать, что гарантированный срок работы базового блока портативного прибора "Флоратест" равен приблизительно 10 годам, что приемлемо для портативного прибора.

Расчет параметров надежности выносного оптического сенсора портативного прибора "Флоратест" осуществляется согласно методике, которая использована для базового блока. Отличие использования методики состоит в том, что в перечень элементов выносного сенсора входят только элементы, параметры которых задаются в единицах FIT.

Выносной оптический сенсор состоит из 4 светодиодов, фотоприемника и операционного усилителя (см. табл. 4).

Таблица 4. Перечень элементов выносного оптического сенсора прибора "Флоратест"

Название	Описание	Количество
<i>Элементы, параметры надежности которых задаются в единицах FIT</i>		
NSPB500	Светодиод	4
OPT301	Фотоприемник	1
AD820AR	Операционный усилитель	1

Согласно с приведенными данными нами получены следующие результаты расчета параметров надежности выносного сенсора портативного прибора "Флоратест" (см. табл. 5).

Таблица 5. Результаты расчета параметров надежности выносного оптического сенсора прибора "Флоратест"

Закон распределения вероятностей отказов	Доверительная граница			
	60 %		90 %	
	часов	лет	часов	лет
Экспоненциальный	20031934	2286,7	16691668	1905,4
DN	530145	60,5	516645	59,0

Из полученных данных видно, что надежность выносного сенсора в несколько раз выше надежности базового блока. Соответственно за надежность целого прибора можно принять надежность базового блока.

Следовательно, гарантированный срок работы портативного прибора "Флоратест" в составе базового блока и выносного оптического сенсора составляет приблизительно 10 лет.

## Выводы

1. Разработаны и проверены при практических расчетах программные модели расчета параметров надежности типичных резервированных систем, а именно: дублированной и утроенной. С помощью разработанных моделей выполнен расчет параметров надежности четырех типов обычных и резервированных систем (1 тыс., 10 тыс., 20 тыс. и 100 тыс. элементов).

2. С помощью разработанных программных моделей рассчитаны параметры надежности портативного прибора "Флоратест", опытная партия которых создана в Институте кибернетики имени В.М.Глушкова НАН Украины. Доказано, что гарантированный срок работы портативного прибора "Флоратест" в составе базового блока и выносного оптического сенсора равен приблизительно 10 лет, что вполне приемлемо для портативного прибора.

## Благодарности

Эта работа частично профинансирована Болгарским национальным научным фондом в рамках совместного болгарско-украинского проекта **D 002-331 / 19.12.2008** "Разработка распределенных виртуальных лабораторий на основе прогрессивных методов доступа для поддержки проектирования сенсорных систем", а также Министерством образования Украины в рамках украинско-болгарского проекта №: **145 / 23.02.2009** с тем же названием.

## Литература

- [Palagin, 2009] Palagin O., Romanov V., Markov K., Velychko V., Stanchev P., Galeyuka I., Ivanova K., Mitov I. Developing of distributed virtual laboratories for smart sensor system design based on multi-dimensional access method // Classification, forecasting, data mining: International book series "Information Science and Computing". Number 8: Supplement to International Journal "Information Technologies and Knowledge". Volume 3/2009. – 2009. – P. 155–161.
- [Гнеденко, 1965] Гнеденко Б.В., Беляев Ю.К., Соловьев А. Д. Математические методы в теории надежности. – М. : Наука, 1965. – 528 с.
- [Стрельников, 2004] Стрельников В.П. Оценка ресурса изделий электронной техники // Математические машины и системы. – 2004. – № 2. – С. 186–195.
- [Азарсков, 2004] Азарсков В.Н., Стрельников В.П. Надежность систем управления и автоматики : учеб. пособие. – К. : НАУ, 2004. – 164 с.
- [Стрельников, 2002] Стрельников В.П., Федухин А.В. Оценка и прогнозирование надежности электронных элементов и систем – К. : Логос, 2002. – 486 с.
- [Romanov, 2007] Romanov V., Fedak V., Galeyuka I., Sarakhan Ye., Skrypnyk O. Portable Fluorometer for Express-Diagnostics of Photosynthesis: Principles of Operation and Results of Experimental Researches // Proceeding of the 4th IEEE Workshop on "Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications", IDAACS'2007. – Dortmund, Germany. – 2007, September 6–8. – P. 570–573.
- [Palagin, 2010] Palagin O., Stanchev P., Romanov V., Markov K., Galeyuka I., Velychko V., Kovyriova O., Galeyuka O., Mitov I., Ivanova K. Calculating of reliability parameters of microelectronic components and devices by means of virtual laboratory // in book "New trends in information technologies". – Sofia: ITHEA. – 2010. – P. 134–143.

## Информация об авторах



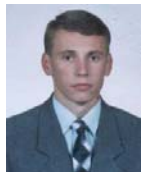
**Александр Палагин** – заместитель директора, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, Академик Национальной академии наук Украины, доктор технических наук, профессор; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: palagin\_a@ukr.net



**Петер Станчев** – профессор, Kettering University, Flint, Michigan (MI), 48504, USA; Институт математики и информатики Болгарской академии наук; ул. Академика Г.Бончева, 8, София-1113, Болгария; e-mail: pstanche@kettering.edu



**Владимир Романов** – заведующий отделом, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, доктор технических наук; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: dept230@insyg.kiev.ua, VRomanov@i.ua



**Игорь Галелюка** – старший научный сотрудник, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, кандидат технических наук; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: galib@gala.net



**Крассимир Марков** – старший научный сотрудник, Институт математики и информатики Болгарской академии наук; ул. Академика Г.Бончева, 8, София-1113, Болгария; e-mail: markov@foibg.com



**Виталий Величко** – докторант, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, кандидат технических наук, доцент; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: [glad@aduis.kiev.ua](mailto:glad@aduis.kiev.ua)



**Александра Ковырëва** – научный сотрудник, Институт кибернетики имени В.М. Глушкова Национальной академии наук, проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: [alexandara.skripka@gmail.com](mailto:alexandara.skripka@gmail.com)



**Крассимира Иванова** – научный сотрудник, Институт математики и информатики Болгарской академии наук; ул. Академика Г.Бончева, 8, София-1113, Болгария; e-mail: [ivanova@foibg.com](mailto:ivanova@foibg.com)



**Илия Митов** – докторант, Институт математики и информатики Болгарской академии наук; ул. Академика Г.Бончева, 8, София-1113, Болгария; e-mail: [mitov@foibg.com](mailto:mitov@foibg.com)

## ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ ДЛЯ ЭФФЕКТИВНОГО ВОСПРОИЗВЕДЕНИЯ ДАКТИЛЬНОГО ЖЕСТОВОГО ЯЗЫКА

Юрий Крак, Юрий Кривонос, Богдан Троценко

**Abstract:** *Complex information technology for visualization communications gesture and mimicry is created.*

**Keywords:** *modelling, sign language, computer system.*

**ACM Classification Keywords:** *I.2.8 Problem Solving, And Search H.1.1 Systems and Information*

---

### Введение и постановка задачи

С созданием независимого украинского государства и признанием на государственном уровне жестового языка как средства межличностного общения и средства обучения глухих лиц возникла насущная необходимость в изучении национального жестового языка и создании современных словарей и средств обучения этому языку.

Языкам свойственны различные диалекты, жестовый язык не является исключением. Жестовый язык, как и любой другой, имеет различные диалекты. Жестовый язык редко используется на телеканалах, к тому же в отличие от устной речи не имеет распространенного письменного эквивалента, как результат, он имеет диалектические различия даже в пределах области. Наличие некоторого «эталона» жестовых единиц (дактилем) в записи позволит облегчить проблему диалектов.

Сейчас жестовый язык можно изучать с помощью учителя, книги или видеозаписи. Но каждый учитель имеет «свой» дактильный жестовый язык, его аудитория обучения ограничена, учитель как «медиа носитель» не копируется, нельзя повторить / увидеть жест вне учебы. При обучении с помощью книги [1]: двумерное изображение не передает динамики и деталей жестовых единиц, книга не является интерактивным медиа-носителем. Даже более современный способ представления жеста с помощью видеозображения [2] имеет свои недостатки: большой объем данных, невозможность рассмотреть жест с разных сторон, кроме того, из отдельно записанных жестов нельзя составить предложение, которое бы выглядело естественно, так как видео не является интерактивным медиа-носителем.

Современное развитие компьютерной техники делает возможным использование более эффективного подхода к обучению жестовому языку, в частности дактильному жестовому языку, который рассматривается в данной работе. Использование компьютерных технологий будет способствовать уменьшению количества диалектов, благодаря распространению «эталонных» жестов. Более того, легкость копирования и использования программ, а также возможность работы по сети Интернет [3], которая позволяет отображать дактилемы на компьютерах различной мощности, будет способствовать распространению знаний и навыков владения дактильным жестовым языком.

**Постановка задачи:** предложить алгоритм и реализацию оптимального по времени расчета состояния пространственной модели при ее высокой размерности, наличия скелета, близкого к скелету руки человека, и обычных мультимедийных требований к аппаратной части компьютера, рассмотреть алгоритмы расчета поверхности модели с целью выявления наиболее эффективного, предложить метод изображения дактилированного слова на основе отдельных дактилем, для проведения вычислений задействовать все ядра процессора. Под оптимальным по времени расчетом подразумевается время, которое соответствует частоте смены 30 и более кадров в секунду.

---

### Математическая модель

Для решения задачи используем скелетную модель руки. Скелет руки математически изобразим в виде дерева  $(V, E)$ , где  $V$  – множество вершин,  $E$  – множество соединений. Каждая вершина соответствует некоторому сочленению костей (звеньям скелета). Каждый элемент отвечает за связь между звеньями,

при этом звено  $e_i$  назовем старшим по отношению к звену  $e_j$ , а звено  $e_j$  – *дочерним* по отношению к  $e_i$ . Каждое звено может иметь не более одного старшего и произвольное количество дочерних звеньев. Звено, не имеющее соответствующего ему старшего, назовем *основным*.

Через  $P(e)$  будем обозначать старшее по отношению к  $e$  звено, если такое существует, и  $\emptyset$  в противном случае.

Для каждого звена поставим в соответствие упорядоченную пару  $h(e) = (x, \theta)$ , где  $x \in R^3$  положение сочленений звена со старшим в системе координат, связанн с  $P(e)$ . Параметр  $\theta$  задает порядок применения углов поворота для преобразования системы координат, связанной со звеном,  $h(V)$  задает исходное положение руки для математической модели  $(V, E)$ .

Для каждого звена  $e \in V$  поставим в соответствие вектор  $r(e) \in R^3$ , состоящий из углов поворота вокруг осей  $OX, OY, OZ$ . Таким образом, совокупность  $r(V)$  задает положение руки, данную совокупность также будем называть параметрами модели. При наличии отношения порядка между элементами множества вершин  $(V)$ , например, по индексу,  $r(V)$  можно единственным образом представить в виде матрицы размерности  $3 \times |V|$ , где  $i$ -й столбец является вектором  $r(e_i), e_i \leq e_{i+1}$ . Данную матрицу будем называть *матрицей параметров модели*.

Основной операцией в данной математической модели является получение положений точек соединений звеньев на основе параметров модели, а также получение матриц перехода систем координат относительно системы координат, которая соответствует основному звену.

Матрицы перехода систем координат определяются формулой 1:

$$M(e) = \begin{cases} M(P(e)) \cdot f(\theta, M_X^{r(e)}, M_Y^{r(e)}, M_Z^{r(e)}), & P(e) \neq 0 \\ I, & P(e) = 0 \end{cases}, \quad (1)$$

где  $M_X^{r(e)}$  – матрица поворота системы координат вокруг оси  $OX$  на угол, который записан в первой строке вектора  $r(e)$ ,  $M_Y^{r(e)}$  матрица поворота системы координат вокруг оси  $OY$  на угол, который записан во второй строке вектора  $r(e)$ ,  $M_Z^{r(e)}$  – матрица поворота системы координат вокруг оси  $OZ$  на угол, записанный в третьей строке вектора  $r(e)$ ,  $I$  – единичная матрица размерности  $3 \times 3$ ,  $f(\theta, M_1, M_2, M_3)$  – произведение матриц в порядке, определенном  $\theta$ .

Положение звена в глобальной системе координат определяется по формуле 2:

$$p'(e) = x(e) \cdot M(e), \quad (2)$$

где  $p'(e)$  положение звена в глобальной системе координат для состояния системы  $M$

На основе данной математической модели построим информационную модель руки, состоящую из математической модели, матрицы точек в пространстве для модели в начальном положении, текстуры поверхности руки, матрицы текстурных координат, матрицы нормалей.

Для получения матриц параметров модели была использована технология Motion Capture (захват движения). Данная технология предусматривает использование оцифрованных данных о движении, записанных с человека-носителя языка жестов. Для этого существует специальная перчатка, имеющая яркие зоны в областях, близких к соединению костей. На несколько видеокамер записывается «проговаривание» всей азбуки дактильного жестового языка. С помощью специального программного обеспечения, Motion Builder, получается динамика изменения положений в пространстве выделенных зон, а на основе этих данных - информация об изменении углов.

Таким образом, за счет использования данных такой природы, нет необходимости во введении ограничений на множество допустимых углов в математической модели. Данные, полученные с помощью

технологии Motion Capture, более реалистично воспроизводят оригинальную динамику «проговаривания» дактилем, чем данные, полученные с помощью дизайнерского программного обеспечения.

### **Алгоритмы расчета точек поверхности руки (скининг)**

Поверхность модели руки представляет собой множество точек, положение которых зависит от положения звеньев. Каждой точке сопоставлено множество  $\{(e_i, w_i)\}$ , где  $e_i$  звено, влияющее на точку,  $w_i \in [0,1]$  - коэффициент зависимости,  $\sum_i w_i = 1$ . Соответственно, положение точки в пространстве определяется формулой  $\sum_i p'(e_i) \cdot w_i$ . Алгоритм расчета данной формулы для всех точек информационной модели и подготовка соответствующих структур данных называются *скинингом* [4].

Для эффективной визуализаций модели на практике было предложено несколько алгоритмов скининга с целью дальнейшей апробации. Все алгоритмы используют массив входных данных точек в исходном положении, записывают результат в массив точек для конечного положения. Типично, что массив результата инициализируется точками с координатами (0,0,0), в процессе работы алгоритма в соответствующие элементы добавляются составные части; таким образом, лишь в конце работы алгоритма результирующий массив содержит правильные положения точек.

Последовательный алгоритм скининга использует представления зависимостей в виде списков (звено; массив коэффициентов зависимостей, каждый элемент которого соответствует элементу из массива координат точек в исходном положении). Он описывается следующим псевдо-кодом:

**для каждого звена  $i$ :**

**для каждой вершины  $j$ :**

**получить координату вершины  $j$  в исходном положении,  
умножить на матрицу перехода в систему координат,  
связанную со звеном  $i$ ,  
умножить на коэффициент зависимости,  
добавить координату к элементу массива результатов**

Алгоритм рассчитан на то, что локальные части массивов координат начального положения, коэффициентов зависимостей, конечных координат на практике находятся в кэш-памяти процессора, что позволит быструю обработку данных с относительно малым количеством операций обмена данными с оперативной памятью.

*Последовательный алгоритм скининга с минимаксной оптимизацией* является улучшенным вариантом последовательного алгоритма скининга, так как на этапе предварительной обработки для каждого звена вычисляются минимальный и максимальный индексы вершин, на которые это звено влияет. Это ориентировано на то, что зависимые точки занимают в массиве координат примерно соседние индексы, что позволит уменьшить количество операций с вершинами, коэффициент зависимостей которых 0.

*Сегментационный алгоритм* скининга является улучшенным вариантом последовательного алгоритма, поскольку он использует представление в виде списков (звено; список пар (начальный индекс, массив коэффициентов зависимостей)). Представление выбирается таким образом, что длина массива коэффициентов зависимостей не превышает заданную величину, например, 256. Каждый элемент с индексом  $i$  из массива коэффициентов зависимостей соответствует элементу  $i + \text{начальный индекс}$ . Алгоритм рассчитан на оптимальное использование кэш-памяти процессора в практическом применении. Он описывается следующим псевдо-кодом:

**для каждого звена  $i$ :**

**для каждой пары из списка**

**(начальный индекс, массив коэффициентов зависимостей),**

**что соответствует звену  $i$ :**

**для каждого коэффициента зависимостей из массива**

**с индексом  $k$ :**

**получить координату вершины в начальном положении**



*с индексом (k + начальный индекс),  
умножить на матрицу перехода в систему координат, связанную со звеном i,  
умножить на коэффициент зависимости,  
добавить координату к элементу массива результатов  
с индексом (k + начальный индекс),*

Прямой алгоритм использует представления зависимостей в виде списков пар (звено; массив пар (индекс вершины; коэффициент зависимости)). Он использует рекурсивную процедуру обхода вершины, что описывается следующим псевдо-кодом. Сам алгоритм заключается в выполнении процедуры обхода старшей вершины.

**процедура Обход вершины (звено i)**

*для каждой пары из массива, соответствующего звену i:*

*получить координату вершины*

*в начальном положении по индексу,*

*умножить на матрицу перехода в систему координат,*

*связанную звеном i,*

*умножить на коэффициент зависимости,*

*добавить координату к элементу массива результатов*

*для каждого звена j, которое является дочерним к звену i:*

*выполнить обход вершины j*

Среди приведенных алгоритмов, на практике прямой алгоритм выполняет минимальное количество операций. В случае, когда сумма объемов памяти, которую занимают точки модели, массив координат начальных и конечных положений, дополнительные структуры данных, меньше, чем объем кэш-памяти процессора, данный алгоритм показывает наивысшую эффективность. В противном случае, рекомендуется использовать сегментационный алгоритм.

Данные алгоритмы были имплементированы, для них были проведены сравнительные тесты на эффективность. Под эффективностью алгоритма подразумевается скорость расчета отдельного кадра, время на подготовку структур данных игнорируется. Тестирование проводилось на скелетной модели, состоящей из 84 460 точек на компьютере с процессором Intel Core2 Duo E8200 (2,66 ГГц), 2 Гб оперативной памяти. Расчет проводился в один поток. Результаты тестирования приведены на рисунке 1.

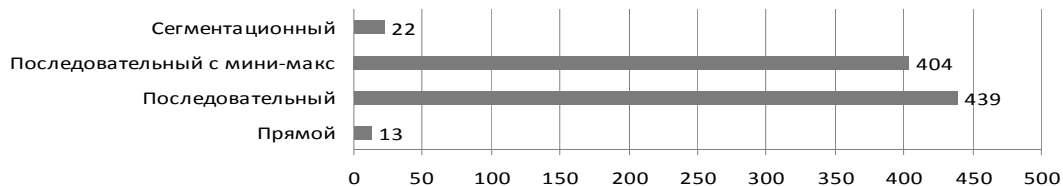


Рис. 1. Результаты тестирования алгоритмов скининга

### Алгоритм дактилирования слова как конкатенация отдельных дактилем

Важным преимуществом предложенной системы является возможность дактилировать произвольные слова, состоящие из отдельных дактилем («букв»). При наличии информации об отдельных дактилемах важно иметь способ их визуальной конкатенации. С использованием модели руки как основы представления графической информации, это становится алгоритмически возможным, в отличие от данных в формате видео.

Была создана база данных, содержащая последовательности состояний модели для каждой буквы. Каждая такая последовательность начинается и заканчивается в некотором «нейтральном» состоянии. Для правильной конкатенации необходимым является подход для определения предельных

состояний каждой дактилемы, где заканчивается информация, которая важна для передачи содержания дактилемы, а также алгоритм создания промежуточных состояний для перехода между дактилемами.

Учитывая природу полученных данных, предельные состояния были установлены экспериментально. Генерирование промежуточных состояний может быть либо алгоритмическим (на основе состояний, которые конкатенируются) или предопределенным. Алгоритмический подход является более удобным, поскольку уменьшает количество необходимой информации, которую нужно подготовить дизайнеру. Для большинства пар дактилем можно алгоритмически сгенерировать промежуточные слайды, например, с помощью линейной аппроксимации на основе предельных состояний смежных букв.

Есть последовательности, к которым применить алгоритмический подход невозможно, например, при дактилировании «Н» - «Г» [1] нужно сначала согнуть пальцы, а потом вращать ладонь. Для этого случая предполагается использование отдельной последовательности состояний перехода.

### Распараллеливание вычислений в задаче подготовки кадра

Большинство выпускаемых современных процессоров оснащено двумя или более ядрами. Дальнейшее развитие ориентировано именно на увеличение количества ядер, чем на наращивание мощности отдельного ядра [5].

Для подзадачи подготовки кадра в задаче отображения трехмерной анимации дактильной азбуки жестов был рассмотрен алгоритм распараллеливания вычислений, поскольку по своей сути подготовка одного кадра является функцией от модели руки и параметров модели, т.е. не зависит от результата вычислений других кадров и от момента вычисления. Такой подход имеет место, когда вычисление нового кадра является ресурсоемкой задачей, например, проведение скининга на процессоре. Для этого был разработан алгоритм, состоящий из  $2 + n$  потоков, где  $n$  количество ядер процессора: основной поток, поток анимации,  $n$  потоков вычисления кадров. Используется тот факт, что вычисления различных кадров является равносильной задачей в терминах процессорного времени независимо от параметров модели. Поэтому вычисление  $k$  кадров можно разбить на  $n$  потоков максимально равномерно таким способом: каждый  $i$ -ый поток обрабатывает кадры с тем номером, остаток от деления которого на  $n$  равен  $i$ .

Основной поток (поток интерфейса) инициирует создание  $n$  потоков вычислений и потока анимации. Поток анимации показывает с заданным промежутком подготовленные кадры, освобождает память после их показа. Поскольку на момент начала анимации должна быть уверенность, что каждый кадр будет готов перед его показом, то поток информации запускается приостановленным, а его запуск делегируется на один из потоков вычисления кадров, в частности, на поток с индексом 0. В дальнейшем, этот поток будем называть *ведущим*.

Каждый поток вычисления кадров в начале получает остаток от деления и начинает цикл вычислений. Также проводится обновление счетчика вычисленных кадров.

Ведущий поток также фиксирует момент времени начала вычислений. После того как было вычислено 5 кадров, он аппроксимирует момент времени окончания вычислений всех кадров. Поскольку длительность анимации известна изначально, то возможно определить, будет ли подготовка кадров выполнена вовремя для непрерывного отображения анимации. Когда аппроксимированного времени до конца завершения вычислений становится меньше продолжительности анимации, ведущий поток запускает поток анимации и прекращает прогнозирование завершения.

Задача распределения потоков на каждое ядро выполняется операционной системой. Но часто бывают ситуации, когда ядра заняты посторонними процессами. Это привело бы к тому, что потоки предобработки получили бы различные кванты процессорного времени, а как следствие, кадры вычислялись неравномерно.

Например, возможна была бы такая ситуация (в примере используется двухъядерный процессор):

№ кадра	0	1	2	3	4	5	6	7	...
Состояние	+	+	+	-	+	-	+	-	...

Где состояние «+» означает, что кадр подготовлен, «-» - не подготовлен.

Для решения данной проблемы был применен подход с использованием семафоров [6,7]: каждый поток проассоциирован одним семафором, который инициализирован единицей. В начале вычисления кадра поток уменьшает свой семафор на единицу, после завершения кадра он увеличивает семафор следующего потока на единицу (последний поток увеличивает семафор нулевого потока). Таким образом, вычисление кадров одного потока не опережает остальные более чем на единицу.

По особенностям своего назначения, семафор не может принимать значения меньше нуля. По своему назначению, значение семафора не может изменяться меньше нуля. Семафор является объектом ядра операционной системы, который отвечает за совместный доступ нескольких потоков к общим ресурсам. У этого объекта две операции – увеличить (на единицу), уменьшить (на единицу). Если значение семафора – ноль, а поток выполняет операцию "уменьшить", выполнение его приостанавливается операционной системой, а значит, становится доступным дополнительное процессорное время. Операционная система ведет учет всех потоков, которые были приостановлены вследствие попытки уменьшить семафор. Когда некоторый другой поток делает операцию "увеличить", то другой поток из списка приостановленных продолжает выполнение.

На практике, такой подход означает следующее: каждый поток анимации в начале работы алгоритма "может" вычислить 1 кадр, за что отвечает значение его семафора. После вычисления своего кадра, он позволит вычисления другому потоку и будет приостановленным, пока поток не увеличит соответствующий семафор. Таким образом, ядра процессора максимально загружены, а кадры вычисляются постепенно.

---

### **Использование технологии внутрипроцессного обмена асинхронными сообщениями**

---

Программа дактилирования украинского жестового языка, как и всякая другая, использует API операционной системы обращения к жесткому диску и ресурсам сети Интернет. Данные функции являются блокирующими, иными словами, вызывающий поток приостанавливается до получения результата.

Стандартный подход к созданию приложений для современных операционных систем (Windows, Linux, Mac OS) предполагает модель событий, как основную концепцию. Различные действия пользователя, действия других программ, либо изменения в системе передаются в программу посредством событий. Для их обработки каждая программа имеет поток, который извлекает события из очереди и вызывает соответствующие методы (обработчики) в программе. Поскольку большинство событий являются извещениями, которые относятся к пользовательскому интерфейсу, данный поток принято именовать потоком пользовательского интерфейса.

В большинстве случаев удается предвидеть среднюю продолжительность обработки API обращения к жесткому диску и выделить отдельные потоки для длительных операций. Исключения составляют те случаи, когда относительно легкие операции ввода-вывода блокируют поток пользовательского интерфейса в результате, к примеру, неготовности жесткого диска. В таких случаях программа не отвечает на другие действия пользователя и выглядит как «зависшая». К сожалению, на практике трудно использовать отдельные потоки для каждой подзадачи, которая требует ввода-вывода.

Для решения этой проблемы предлагается технология обмена сообщениями между разными потоками внутри единого процесса.

Технология предлагает использование шины обмена данными, на которой регистрируются обработчики сообщений. Каждый обработчик имеет свой идентификатор, один или несколько потоков для обработки. Каждое сообщение содержит идентификатор соответственного обработчика, набор параметров, переменные результата, объект ядра операционной системы, типа «событие», для извещения об окончании завершения обработки данного сообщения.

Предлагается реализация данной технологии в среде .Net. Ее диаграмма классов представлена на рисунках 2, 3.

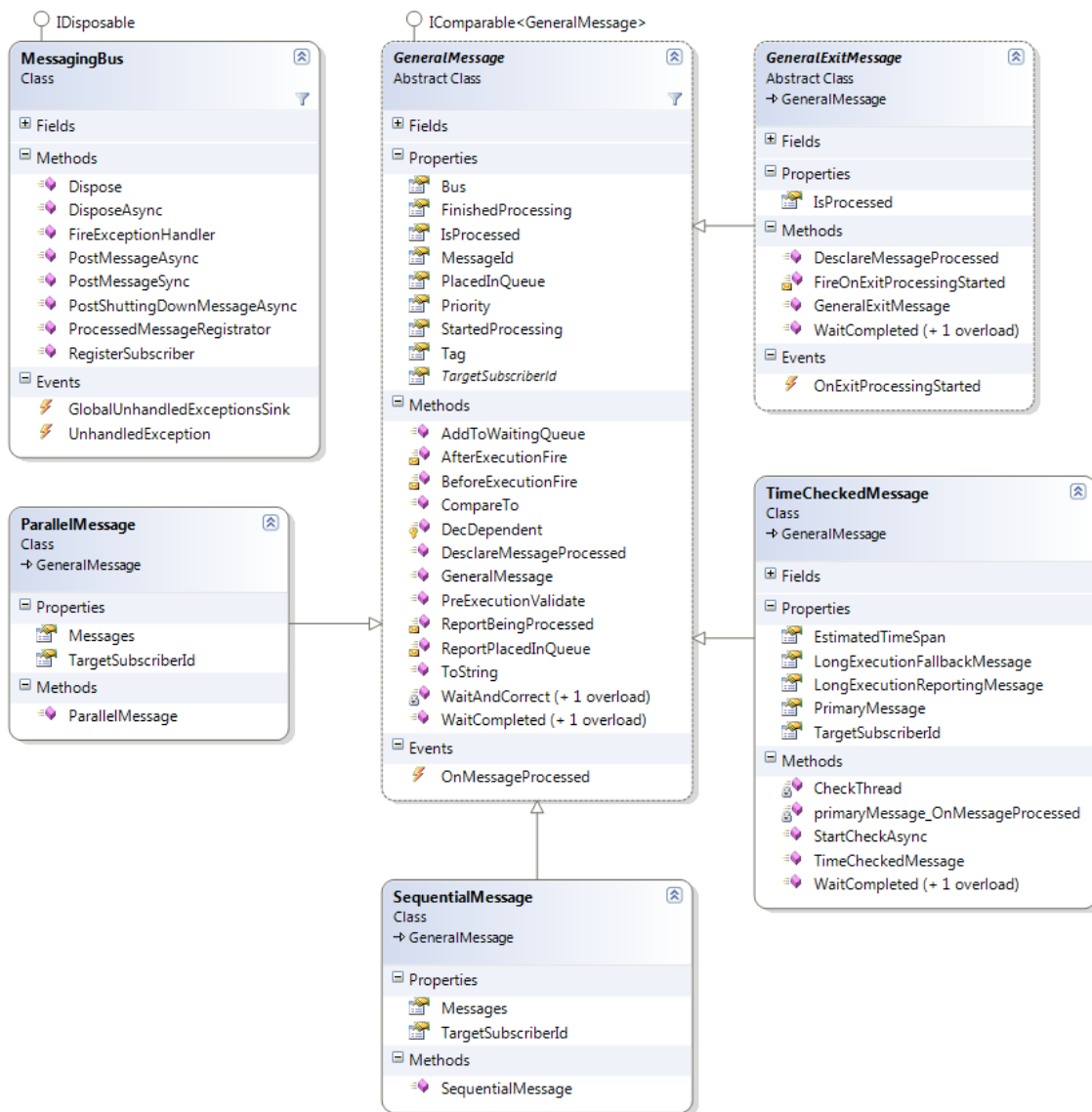


Рис. 2. Диаграммы классов для шины и основных сообщений

Основной класс, `MessagingBus`, описывает шину. С помощью метода `RegisterSubscriber` на шине регистрируются обработчики – классы, которые наследуют `GeneralSubscriber`. Метод `PostMessageAsync` передает сообщение обработчику и возвращает управление вызывающему потоку. Дополняющий его метод `PostMessageSync` дополнительно блокирует вызывающий поток до окончания обработки сообщения.

Абстрактный класс `GeneralMessage` представляет обобщенное сообщение. В частности, он имеет свойства `TargetSubscriberId` – идентификатор обработчика, `Priority` – уровень приоритета; метод `AddToWaitingQueue` регистрирует другие сообщения, которые влияют на результат данного, метод `WaitCompleted` блокирует вызывающий поток до окончания обработки сообщения; событие `OnMessageProcessed` уведомляет об окончании обработки сообщения.

Отдельно выделен класс `GeneralExitMessage`, который предназначен как базовый класс для сообщения, извещающего о завершении работы программы. Он должен быть передан на шину с помощью метода `PostShuttingDownMessageAsync`, который передаст это сообщение обработчику только тогда, когда на шине больше не будет других сообщений в очереди, либо в состоянии обработки.

Класс `SequentialMessage` принимает в качестве параметра массив других сообщений для поочередной обработки. Подобный ему класс `ParallelMessage` также принимает в качестве параметра массив других сообщений, но обработка его заключается в разовой передаче всех сообщений на шину и ожидания окончания их обработки.

Класс `TimeCheckedMessage` позволяет контролировать время обработки сообщения (`PrimaryMessage`); если время превышает заданный интервал, то на шину также передается `LongExecutionReportingMessage`. В случае передачи последнего, также будет передано сообщение `LongExecutionFallbackMessage` по завершению обработки основного сообщения.

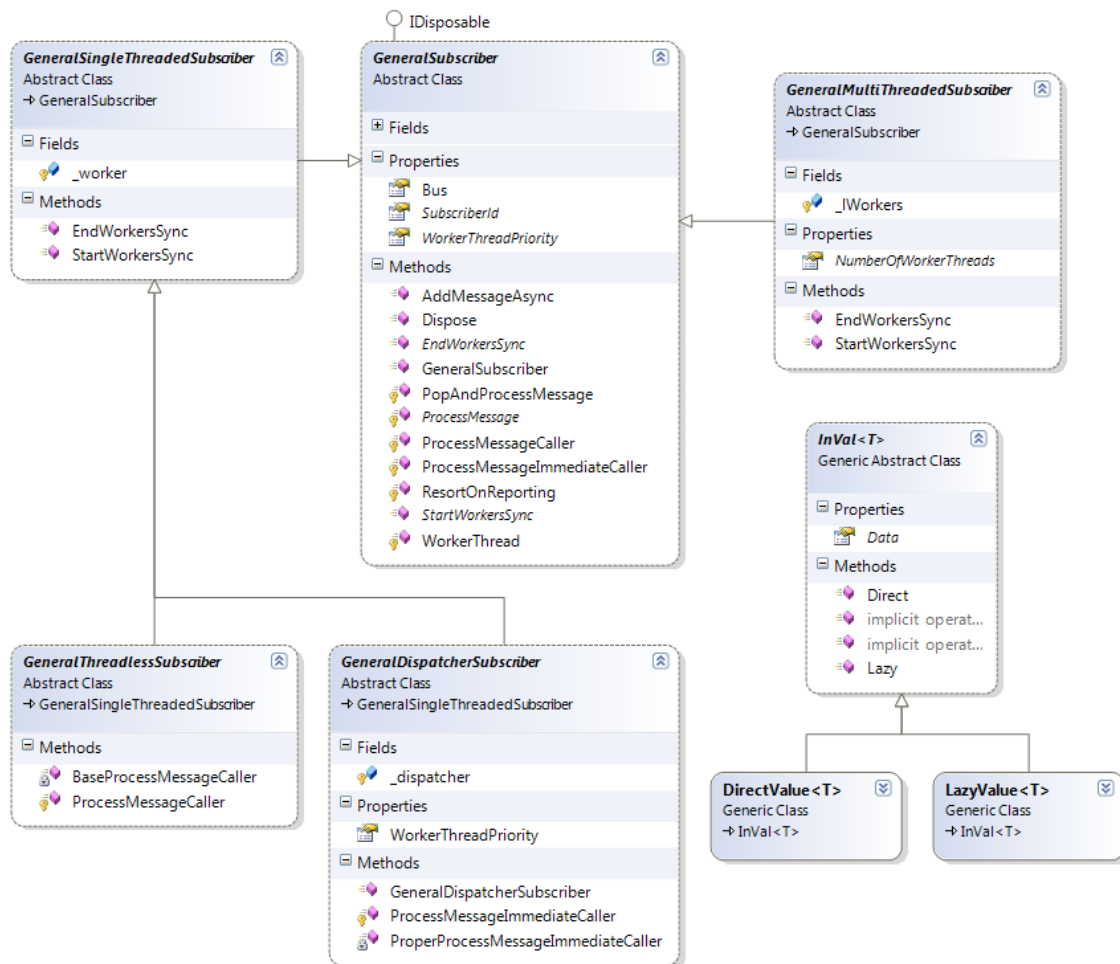


Рис. 3. Диаграммы классов для обработчиков сообщений и входных данных

По своей природе, программирование с использованием сообщений можно назвать отложенным вызовом функций. Поэтому, возможно два сценария: расчет входных параметров происходит 1) в момент организации сообщения, 2) перед началом обработки. Реализация данных сценариев произведена с помощью параметризованных классов `InVal<T>`, `DirectValue<T>`, `LazyValue<T>`. Конструктор `DirectValue` получает непосредственно значение, конструктор `LazyValue` получает делегат на функцию, которая вычислит входной параметр непосредственно перед началом обработки.

Класс `GeneralSubscriber` описывает общие свойства и поведение обработчика, реализует методы для работы с очередью сообщений. Класс `GeneralSingleThreadedSubscriber` позволяет осуществлять работу в один поток, `GeneralMultiThreadedSubscriber` – в несколько; число потоков регулируется свойством `NumberOfWorkerThreads`. Класс `GeneralDispatcherSubscriber` выполняет работу в потоке пользовательского интерфейса.

Использование данной технологии позволяет удобно описывать рабочий цикл программы. Пример использования приведен в листинге 1.

```

Bus.PostMessageAsync(new SequentialMessage(
    new ParallelMessage(
        new SequentialMessage(
            new ShowStatusMessage() { Status = "Загрузка модели..."
        },
        new ComputingLoadModelMessage()
        {
            Path = "FinalData",
            ModelConfiguration = ModelConfiguration
        },
        new ShowStatusMessage() { Status = "Модель загружена." }
    ),
    new ParallelMessage(LoadCharacterMessages.ToArray())
),
new ConfigureModelMessage(),
new PositionHandMessage(),
new ShowStatusMessage() { Status = "Инициализация завершена." },
new EnableUsageMessage()
));

```

Листинг 1. Пример использования технологии передачи асинхронных сообщений

---

## Использование технологии передачи данных через сеть Интернет

---

Для распространения программного приложения, позволяющего дактилировать отдельные буквы и слова, использовано обобщенный способ оптимальной передачи динамично созданного изображения в сети Интернет с возможностью управления им [4]. Поскольку вычислительные возможности компьютеров широкого круга пользователей значительно отличаются, также как и возможности их каналов связи, данный аспект был учтен при продумывании способа массового обслуживания, посредством передачи данных через сеть Internet. Разработанная технология имеет ряд преимуществ над уже существующими, которые или не имеют интерактивности, или для ее достижения необходимо использовать специальные плагины, которые необходимо установить на компьютер клиента.

Технология передачи данных содержит трехуровневую информационную модель коммуникации между серверной и клиентскими частями: на первом (высшем) логическом уровне видеoinформация передается из серверной стороны на сторону клиента. На следующем втором логическом уровне, сервер кодирует видеoinформацию, фрейм за фреймом, клиент проводит инициализацию (предусматриваемую сервером), а также воспроизводит картинку на основе команд. На этом логическом уровне программист (инженер, разработчик) может учитывать специфику задачи и структуру медиаданных, т.е. предоставить собственную пару компрессора-декомпрессора данных. На этом уровне отдельный поток отвечает за обслуживание именных каналов и передачу команд компрессора. Рекомендуемый уровень компрессии выходит из расчета, что 10 процентов времени канал должен либо простаивать, либо резервироваться на непредвиденную ситуацию, например, передачу данных другим программным приложениям на стороне клиента. На следующем третьем логическом уровне команды сериализуются на стороне сервера и десериализуются на стороне клиента. Это последний логический уровень данной технологии. Информация на данном уровне передается с использованием HTTP (Silverlight / Asp.net).

---

## Выводы

В данной статье впервые для воспроизведения дактильных украинских жестов использовано анимацию трехмерной модели руки, рассмотрены алгоритмы расчета поверхности руки с помощью трехмерной модели руки, базирующейся на скелетной модели, выполнена реализация этих алгоритмов и практическая апробация, рассмотрен алгоритм анимации дактилирования слова на основе данных об отдельных дактилемах, усовершенствованы алгоритмы подготовки кадра для воспроизведения процесса скелетной анимации дактильного жестового языка за счет поддержки многоядерных процессоров. Данные алгоритмы и методы совместимы с технологией отображения информации средствами Интернет для воспроизведения процесса анимации дактильного жестового языка, что позволит создать соответствующее программное приложение.

Дальнейшие исследования будут направлены на повышение качества отображения руки, улучшение алгоритмов отображения с целью поддержки использования программы на компьютерах с меньшей мощностью.

---

## Благодарности

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

---

## Библиография

- [Кульбіда С. В.] Кульбіда С. В. Українська дактилологія. К. : Педагогічна думка, 2007. — 255 с. — ISBN 978-966-644-076-4.
- [Воскресенский А. Л.] Воскресенский А. Л., Халагин Г. К. От звучащей речи к жестовой // Речевые технологии. — 2009. — № 1. — С. 99—106.
- [Бармак О. В.] Бармак О. В., Крак Ю. В., Троценко Б. А. Технологія оптимізованої передачі жестової мови в мережі інтернет // Проблеми програмування. — 2009. — № 3. — С. 73—80.
- [Domine S.] Domine. S. Mesh Skinning — Режим доступа: <http://developer.nvidia.com/attach/6662>. — Название с экрана.
- [Merritt R.] Merritt R. CPU designers debate multi-core future — Режим доступа: <http://www.eetimes.com/showArticle.jhtml?articleID=206105179>. — Название с экрана.
- [Agarwal A.] Anant Agarwal, Srinivas Devadas, Henry Hoffmann Partitioning Strategies for Concurrent Programming // Technical Report MIT-CSAIL-TR-2009-026. — 2009. — Режим доступа: <http://dspace.mit.edu/bitstream/handle/1721.1/45567/MIT-CSAIL-TR-2009-026.pdf>. — Название с экрана.
- [Massingill B.] B. Massingill, T. Mattson, B. Sanders. Patterns for parallel programming — Reading, Massachusetts : Addison-Wesley Professional, 2004. — 384 p.

---

## Authors' Information



**Lurii Krak** – senior scientist, V.M.Glushkov Cybernetics Institute of NASU,  
40 Glushkov ave., Kiev, Ukraine, 03680, [krak@unicyb.kiev.ua](mailto:krak@unicyb.kiev.ua)



**Lurii Kryvonos** - Director Deputy, V.M.Glushkov Cybernetics Institute of NASU,  
40 Glushkov ave., Kiev, Ukraine, 03680



**Bohdan Trotsenko** – junior scientist, V.M.Glushkov Cybernetics Institute of NASU,  
40 Glushkov ave., Kiev, Ukraine, 03680 [bohdan@trotsenko.com.ua](mailto:bohdan@trotsenko.com.ua)

## A GRAPH MODEL OF SHOCKED FINANCIAL SYSTEM

Velina Slavova, Lyubomir Mirchev, Hristian Elefteridis

**Abstract:** We applied mathematical modeling in order to understand the behavior of the banking system as a network infrastructure. Based on graph modeling, we developed program modules for simulating the propagation of a financial shock over a system of interconnected banks. The simulation imitates a wave of contagion of the banking system, which is generated as a random graph. In order to analyze the behavior of the system in general, we perform a multiple generation of random graphs and consider the obtained mean values. We analyze the influence of a set of input economic parameters with regards of the properties of the banking system as a "critical infrastructure". We used different subordination models of the graph structure, with regards of contemporary financial theories and taking into account recent documents of the European financial commissions. The results of the simulations permit to do important recommendations concerning the methods for protecting and optimizing the banking system.

**ACM Classification Keywords:** I.6.5: Model Development, I.6.6: Simulation Output Analysis. G.2.2 Graph Theory, G.2.3 Applications

---

### INTRODUCTION

The recent global experience highlights the need to monitor systemic risks arising from both the macroeconomic developments in the economy and from the global financial markets. This requires new concepts, methodologies and models for financial system transformation and network crisis management. We applied mathematical modeling in order to understand the behavior of the banking system as a network infrastructure.

The mathematical formalism in this work is based on graphs. The nodes are banks and the edges imitate the interbank connections. We did four main successive steps in our approach. The first was to elaborate algorithms running on the graph in order to adjust the model from the point of view of its economic parameters and to simulate the behavior of banks system when it is submitted to an external economic shock. The second step was to include in the system additional processes in order to better imitate some particularities of the financial market in situation of crises. The third step was to adjust the structure of the underlying graph as self-organized dynamical system. On this bases, we propose possible ways to protect the system, namely to limit the propagation of the shock. We performed computer simulations of the proposed measures of protection. The rest of this paper is organized following these main steps.

**Underlying economical calculus:** We build a random graph with a chosen number  $N$  of nodes (banks) which have lent to one another with probability of  $p$ . The probability  $p$  expresses an important economic parameter - the connectivity of the banking system. So,  $p_{ij}$  is the probability that bank  $i$  has lent to bank  $j$ . The existence of an interbank borrowing from bank  $i$  to bank  $j$  is expressed by means of an edge between these two banks (the graph is oriented). The generation of the graph is performed using a simple random selection of  $Z$  edges from the full set of  $N*(N-1)$  edges of the corresponding complete graph.  $Z = N*(N-1)*p$ , where  $p$  varies from 0 to 100%.

We construct the graph of a hypothetical banking system by using the following **inputs**:

$N$  – number of banks;

$p$  – probability of connection;

$E$  – total external assets of the banking system;

$E/A$  – percentage of system's external assets to system's total assets;

$c/a$  – percentage of bank's capital to total bank's assets.



Each node is generated with its initial balance<sup>1</sup>, following a step-by step repartition of the actives and passives of the entire system. In big, the parameters of the balances - capital, assets and deposits are calculated depending on the number of connections that each node has obtained after the generation of the graph. More precisely, the banks balance sheets are calculated by starting with the external assets of the banking system as a whole ( $E$ ). Having the formula for the total assets  $A = E + I$ , where  $I$  is the total interbank assets in the system and having  $E$  and  $E/A$  as an input, we find  $A$  and  $I$ :  $A = E / (E/A)$  and  $I = A - E$ . Knowing the total size of the interbank exposures, the **weight  $w$  of each edge** is found by dividing  $I$  to the total number of edges:  $w = I / Z$ .

From there we find the individual bank's interbank assets and borrowings by multiplying  $w$  to the number of outgoing and incoming connections to each node (bank) in the graph. We denote the individual bank assets by  $a$ . So:  $a_i = e_i + i_i$ , where  $e$  are the external assets i.e. loans and other investments to non-bank counterparties, and  $i$  denotes the interbank assets, i.e. exposures to other banks in the system. On the other side of the balance sheet we have the liabilities, denoted by  $l$ , so  $l_i = c_i + d_i + b_i$ , where  $c_i$  is the capital of bank  $i$ ,  $d_i$  are the deposits (from non-bank customers) of bank  $i$ , and  $b_i$  denotes the borrowings (from other banks) of bank  $i$ . As every balance sheet,  $a_i = l_i$ . The algorithm which calculates these bank-balance parameters has one iterative loop for distributing the "rests of the total money" after this first distribution.

The interbank assets  $i$  of one bank are the borrowings  $b$  of another – these linkages are used as a **shock transmitting channel**.

## THE MODEL OF THE SHOCK

We simulate a propagation of damages over the nodes of the graph. The edges are channels of transmission. We concenter that the initial shock  $S$  arises in the economic environment, and, being external for the system, it affects the external assets  $e$  of a bank. The size of the shock  $S$  is a chosen percentage of the external assets of the initially shocked bank. In the proposed simulation, when a bank receives a damage  $s_i$ , its capital  $c_i$  is the first to absorb the losses (figure 1).

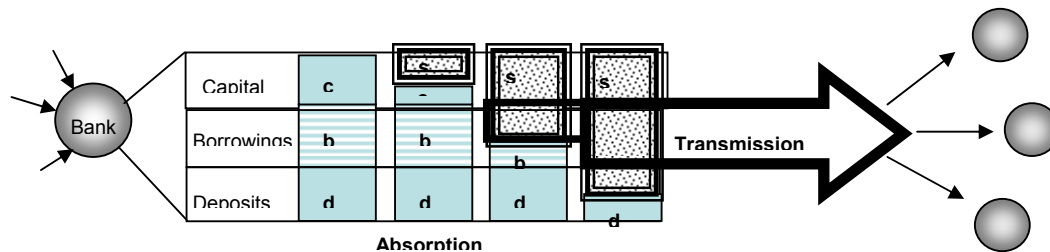


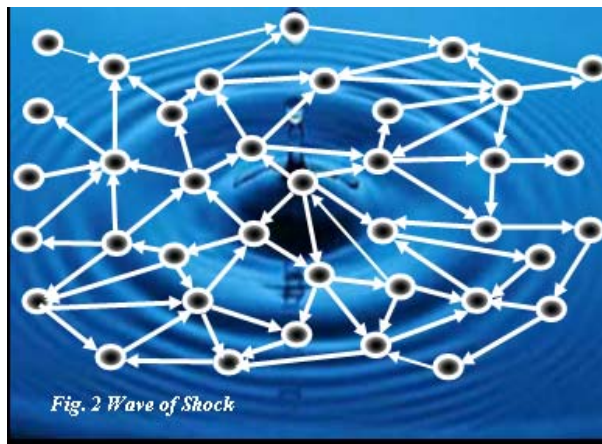
Figure 1. Shock absorption and transmission.

If the damage is greater than the bank's capital, then the bank is in failure. The borrowings  $b_i$  are the next to absorb the losses, thus the shock is transmitted to the net. The damage is then transmitted to the network, harming the neighbor bank-nodes. So, in the process of propagation we distinguish two states of a bank – "touched bank", when the bank settles down its capital, but resists the shock and "bank – transmitter", when the bank has lost its capital and the damage has touched its borrowings from other banks. If the losses exceed the borrowings  $b_i$ , then the clients' deposits  $d_i$  are damaged. We called that third possible state of a node a "dead bank". The economic theory does not take into account the damages received by the deposits (the non-bank clients) and our analyses is not centered on the harmed deposits.

As illustrated on figure 1, the initial shock absorber of the damage  $S$  is the capital  $c$ , the transition of the damage to the system is via the borrowings  $b$  and the final shock absorber of the damage are the deposits  $d$ .

<sup>1</sup> The precise procedure is described in Mirchev, Slavova, Elefteridis 2010.

The damage, transmitted to each of the bank-borrowers depends on the weight  $w$  of the connecting edge. In the model, the edges have the same weight and we distribute the unabsorbed damage equally on each outgoing edge. The attained banks receive the damage in the same way – they decrease their capital, further they lose their borrowings and become transmitters of the contagion wave. The process of propagation goes on until there are not more touched banks (the shock is absorbed by the system).



One may do an analogy with the propagation of a wave in a fluid as it is illustrated on fig. 2. Each node has two functions: 1. the bank is an *absorber* of damages by means of its capital and 2. the bank is a *transmitter* of the shock after the threshold of its capital is reached. The number of edges included in the process has also a double function. The distribution of the damage over the outgoing edges diminishes the fractions of damage and increases the number of touched banks (*diversification*). There is one more phenomenon which deserves our attention: the nodes stock the incoming damages via its entering edges. A given bank can initially resist the damage by absorbing it, but after some time can 'brake down' as the propagation of damages has reached it several times via different paths (note also that the graph has cycles). We called this phenomenon a *gradual exhausting*. It causes a postponed rising of the number of failed banks.

The output that we register is the number of failed banks (bank defaults).

The simulation of the system's behavior is done by means of measuring the output when varying the values of the input parameters.

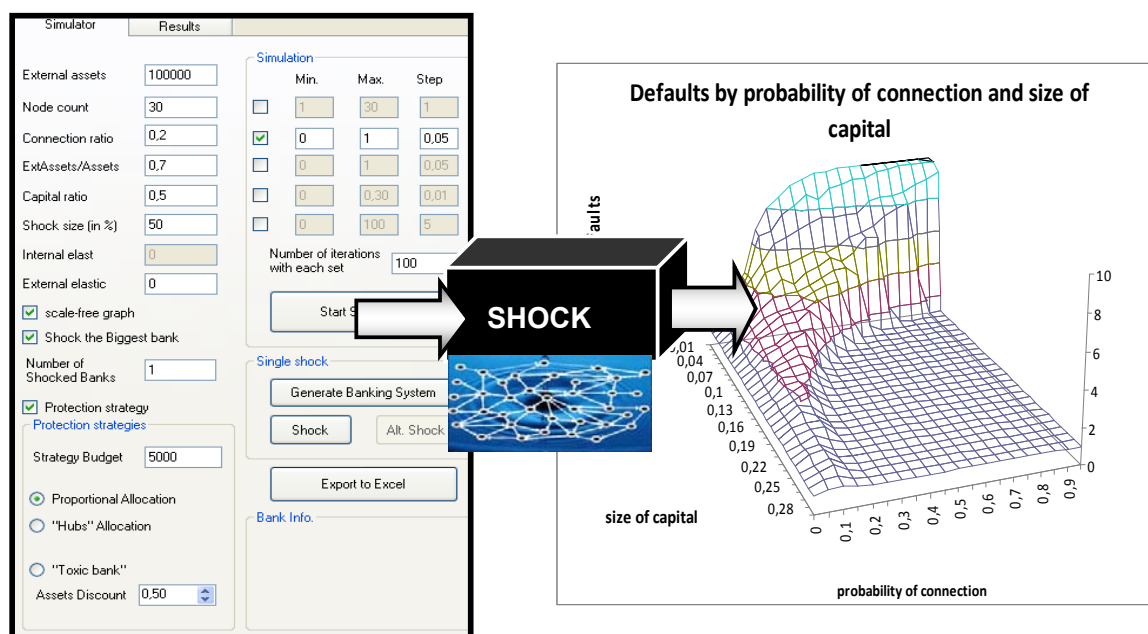


Figure 3. Simulation architecture

The influence of each input parameter is examined separately, by increasing gradually its value in small successive steps, when the other parameters stayed fixed (Figure 3.). For each value of the examined parameter,

we performed multiple generations of random graphs (we performed 100 generation per step), we run the simulation of the shock and we calculated on the bases of the received 100 outputs the *average* of the number of failed banks. The programs are realized in Visual C#. The outputs, transferred in MS Excel tables, are used as source for the graphics and the further analyses of the system's behavior.

We conceived, implemented and tested several algorithms, realizing the described above process, called "Wave". The accuracy tests were performed by checking the total bank balances of the entire system before and after the wave of shock (see table 1). Each point of the graphic-examples given further is obtained after 100 runs of "Wave" on random graphs. That has required a special attention to the complexity of the used algorithms and especially of the one which calculates the propagation of the shock.

## ANALYSES OF THE RESULTS

The described algorithms and simulation method were used for the initial evaluation of the systems behavior. The interface and the outputs of the simulation are given on figure 4 and figure 6 respectively.

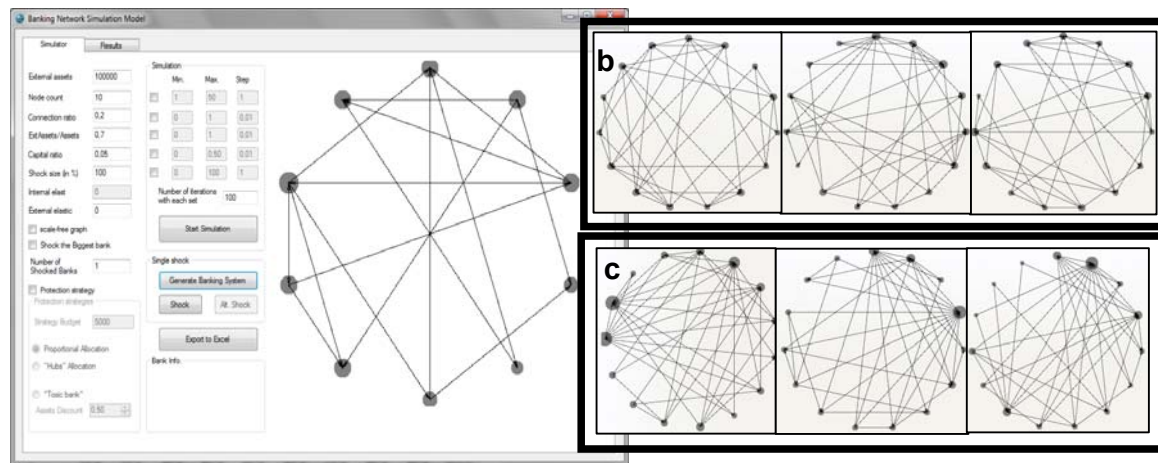


Figure 4. – Simulation model interface – written on Microsoft Visual C# Express; b) Examples of three random graphs with 15 nodes and connection ratio  $p=20\%$ ; c) Examples of three scale-free graphs, connection ratio  $p=20\%$ .

From the point of view of the analogy with a fluid system that we made, the connection ratio  $p$  is a parameter with a crucial importance. In order to check the processes that we have supposed existing in such system, we give the example of analyses of the influence of the connection ratio from a more formal point of view. The results of simulation are shown on figure 5. The simulation is made for values of  $p$  between 0% and 100% - from a disintegrated system to a completely connected system. When the system is disintegrated, only the initially shocked bank defaults. With the rise of connectivity, the number of failed banks increases till a certain point of deterioration. In the interval where the connectivity is very poor, the graphics rises quickly as the damages are distributed in big fractions and the contaminated banks immediately fail harming the deposits. In this case, to suppress the wave of shock, the system utilizes also its second shock absorber.

With the further rise of connectivity, the portions of damage transmitted to the “neighbors” decreases and the number of banks which succeed to cushion the shock rises. That corresponds to the flattened part of the graphics on fig. 5. The effect of gradual exhausting is clearly observable on the graphics, expressed by the second rise of the curve. In fact, with the rise of connectivity, the damages distributed to each of the neighbors decrease and the deposits are less utilized as shock absorbers. The pathways of the contagion wave become longer and start crossing the same nodes several times. So, after a number of “attacks”, all these banks “give up” and fail “en bloque”.

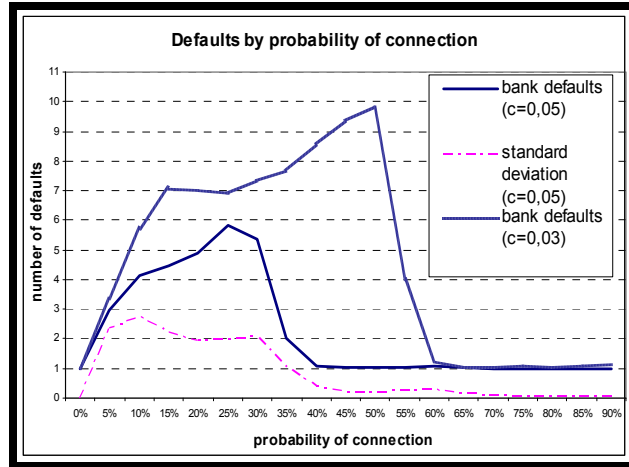


Fig.5. Number of defaults depending on the graph connectivity

After some degree of connectivity, the distribution of the shock becomes big enough to produce fractions which are absorbed immediately from the touched banks. That corresponds to the saturation point of the graphics. The positioning of the point of saturation depends on the capital – the first shock absorber. A system with less capital (on the graphics  $c=0,03$ ) needs more connections to absorb the shock. That is illustrates also on figure 6.d.

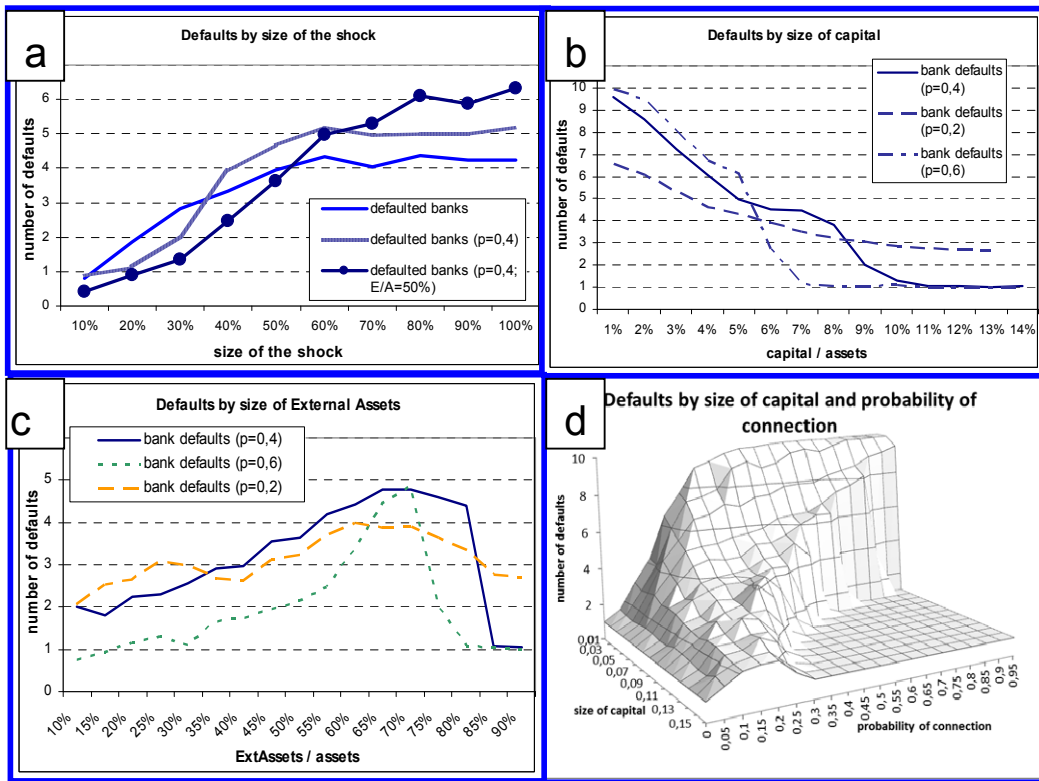


Figure 6. Outputs of the Simulation. Graphics - failed banks depending on a) the size of the shock; b) the size of the system’s capital ; c) the external assets; d) the connection ratio and the capital

The economic interpretation of the influence of the input parameters is deeply discussed in Mirchev, Slavova Elefteridis 2010. Here we provide an example - the analyses of the influence of the size of the shock (fig. 6.a).

While increasing the size of the shock (fig 5.a), the extent of contagion also increases to a certain point where the shock is distributed to enough banks and can be absorbed. This mechanism depends on the level of bank interconnectedness. In more connected banking system the contagion effect is greater. Nevertheless, the contagion pattern is different: while the shock is relatively small, the more connected network is absorbing the shock better as it is distributing small fractions of the shock to many neighbors. After certain point the effect of contagion prevails over the effect of diversification and the contagion sequence rolls over. After a certain point of saturation the increase in the size of the shock is not provoking additional failures because the shock is distributed to sufficiently many nodes in the system.

The simulation model developed using graphs and algorithms for economic parameters calculation and the propagation wave of the shock in the banking system show an adequate behavior in the sense that the simulation results are easily explicable in terms of parameters and financial and economic dependencies. The model reveals that the stability of the system depends not only on the individual bank's stability, but also on the intensity and the size of interbank connections i.e. how integrated the banking market is.

The system's behavior will be different if its resistance decreases during the wave propagation because of the fact that the absorbers are weaken. And, if the shock touches more nodes, the wave propagation will be changed. In fact, that is a more realistic model of the financial system.

### FINANCIAL MARKET REALITIES - WEAKEN ABSORBERS AND MASSIVE SHOCK

Further, instead of the initially modeled perfect market liquidity, we introduced the liquidity effect described by the elasticity of the assets' price to the assets sales and the **mark-to-market accounting** effect to the banks' balance sheet accounts. That financial phenomenon is like a progressive decrease the resistance of the shock absorbers, as it decreases the power of their capital. The assets price is decreasing with the same proportion as the assets sold on the market compared to the total assets in the system. We introduced coefficient '*elasticity*', affecting the magnitude of the liquidity effect. Value 0 means no liquidity effect or perfectly inelastic asset price (perfect liquidity), and 1 means full liquidity effect or unitary elasticity of the asset price.

$$\% \text{ of change in assets price} = 1 - \text{elasticity} * (a_{\text{sold}} / A)$$

Banks' assets are re-evaluated on each simulation cycle with the current assets market price. When a bank defaults it sells all its remaining assets and exerts pressure on the assets prices.

In parallel with that the balance sheets of all other remaining banks are revaluated and the amount of assets is decreased in accordance with the new prices. The effect of the mark-to-market accounting principle is that is weakening all the banks. A kind of additional shock cumulates over the balance sheet. As it was supposed, the contagion profile with mark-to-market accounting resembles to a contagion profile of less capitalized banking system (fig 7.).

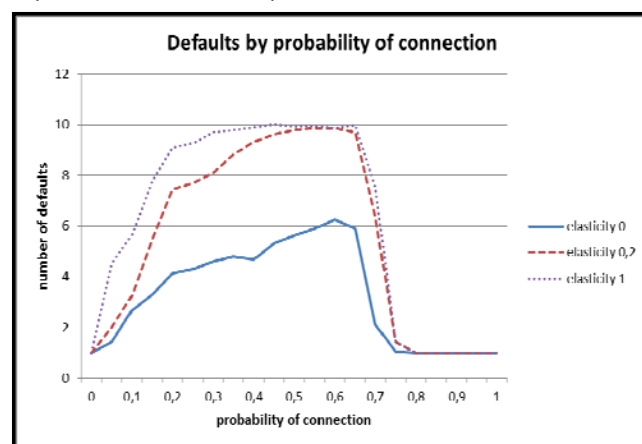


Figure 7. Defaults by connection ratio and liquidity effect

As the plots of the simulation conform (figure 8.), the system loses resistance. With the same level of initial shock the banking system needs more capital to withstand the contagion. The detailed economic analysis of the liquidity effect is given in Mirchev et al 2010.

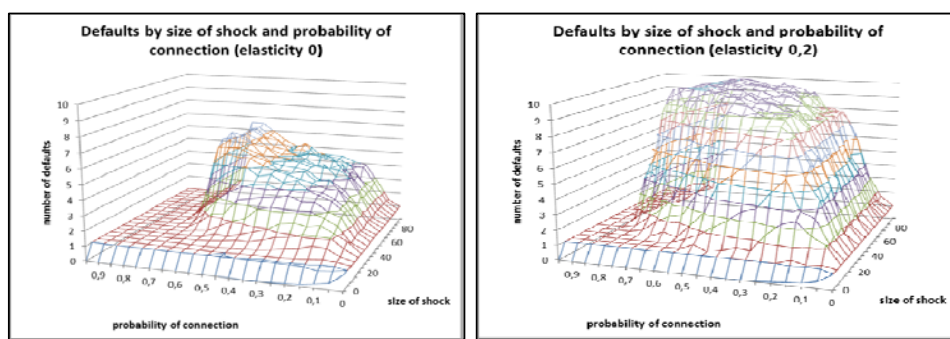


Figure 8. Influence of the liquidity effect – less resistant system.

Most of the external shocks would affect several or all banks simultaneously. We introduced the ability to shock several banks at the beginning of the simulation. The contagion profile shows how vulnerable and fragile a banking system is (fig. 9). When increasing the number of initially shocked banks the contagion effect is stronger (wider contagion area).

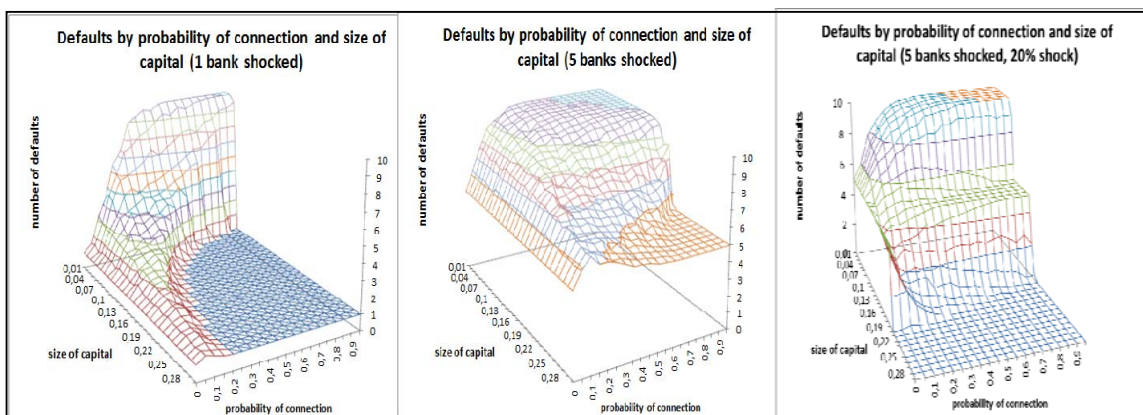


Figure 9. Influence of the Massive chock a) one bank b) 5 banks; c) 5 banks 20% of shock

We conducted similar simulations, but this time maintaining relatively identical size of the shock in the different scenarios i.e. increasing the number of initially shocked banks while decreasing the shock size for each bank (figure 9.c). A slight increase in the contagion area is observable, it is due to the fact that the initial shock is spread to more banks and thus the number of “powerful” contagion channels increases. We can see a slight increase in the contagion area (fig. 9.c.) due to the fact that the initial shock is spread to more banks and thus we have more contagion channels. By further increasing the number of initially shocked banks, the relative size of the shock for each bank is getting smaller and we see a new three staged contagion profile. First, when the banks have lower capital shock absorbers, the contagion effect works and we have high number of defaults. At some point by increasing the banks’ capital the effect of diversification starts prevailing and the number of defaults is limited only to the initially shocked banks. By further increasing the capital level, having in mind that with high number of initially shocked banks the size of the shock is relatively small, the banks are getting able to withstand the shock and the system scores no defaults.

The influence of the liquidity effect when the shock touches several banks can be previewed as gradually weakening the system in addition of the massive shock. That is the worse situation, as the two factors harm the system's resistance (see figure 10). Several simulations of different scenarios are performed using the possibility to vary the input values and the results are analyzed from the economic point of view in details in Mirchev, Slavova Elefteridis, 2010.

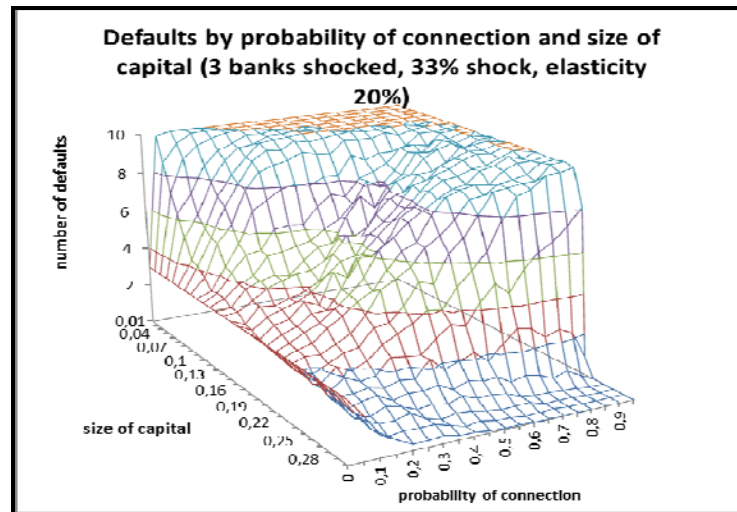


Figure 10. Realistic situation – massive shock + liquidity effect

The simulations gave the possibility to figure out possible strategies to protect the banking system. That was done with regards of the existing modern tendencies in the economic theories and taking into account the recent documents of the European Union. At EU level for *critical infrastructures* have been adopted stock exchanges and settlement systems. Considering the importance of the banking market (which dominates over other sectors of the financial market in the EU), we suggest its consideration as a separate critical infrastructure (social networking infrastructure). It is the economic infrastructure connecting the market participants in the economy and facilitating the processes of financial resources transformation. That assumption has necessitated modifying the underlying graph.

## NEW TOPOLOGY AND PROTECTION STRATEGIES

Until recently, the banking system was considered as a set of financial institutions competing in a specific market – the banking market. In this respect, their role was not considered different from any other market player on the financial and non-financial markets. When a bank fails, the law provides protection to the creditors and in most cases to those entrusted their money to the banks - the depositors. But the crisis has shown us that the disturbance occurred in the financial market is rapidly transmitted to the rest of the economy. All the entities, relying on the banks services for conducting their businesses, are also affected adversely. The social function of the banks comes into focus - their role as financial intermediaries in the economy.

A model for analyzing and strengthening the stability of the banking market, considered as critical infrastructure would include the following steps:

1. Mapping the real topography (banks inter-connectivity) of the banking network.
2. Identification of the hubs in the system – the supervisory efforts could be focused to these nodes depending on their importance.
3. Assessing the extent of the threat / possible damage, which a hub brings to the system. The risk depends on the hub's size and connectivity with other hubs and nodes in the network. This evaluation is performed through simulations or analysis through the development of a "fault-tree" for spreading the initial shock.
4. Budget analysis – determines the optimum allocation of resources. Possible tool is the "network-wide investment" - after assessing the possible negative effects of each hub in the system, the investments

are allocated in such a way that minimizes the overall negative effect. Priority is given to the most important hubs, which have the greatest impact on the network stability.

Many communication and social networks have power-law link distributions, containing a few nodes which have a very high degree and many with low degree. The high connectivity nodes play the important role of hubs in communication and networking, a fact which can be exploited when designing efficient protecting algorithms. The presence of self-organized criticality in social networks is often evidenced by a power-law scaling of event size distributions, which can be measured by linear regression on logarithmic axes. The power-law scaling can also be induced by the stochastic nature of the social phenomena.

Financial markets are largely integrated, but the institutions responsible for their supervision and safeguarding the financial stability remain divided along the national lines. Studies on the stability of financial systems (Allen 2000, Nier 2008) used the assumption that the participants are equal, and the distribution of links is a random. Studies on social networks (as are the banking/financial markets) show that their structure is more complex (Lewis, 2009) and have the characteristics of scale-free networks and small-world networks.

Scale-free networks (fig. 10.b) are networks whose degree distribution follows a **power law** - probability of a node to make connection to other depends on the number of connections, which it owns. The radius of the network - the number of edges between the two most distant nodes, is relatively small. They are clustered - some nodes called "hubs" have much more connections than others. Hubs in the banking system are the systematically important banks.

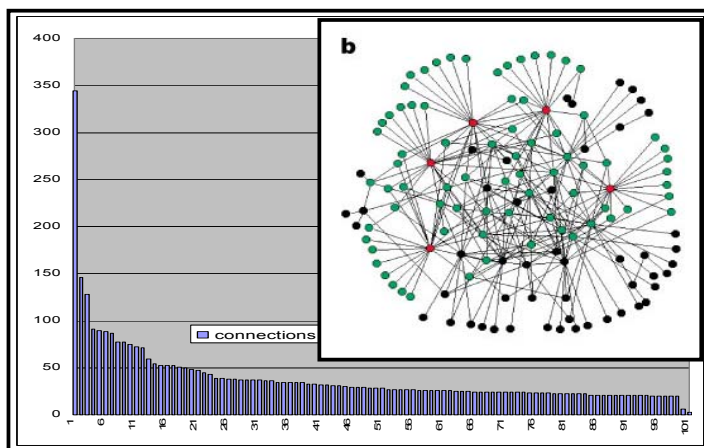


Fig 10. A scale-free graph (b) and the distribution of its connections

We implemented the Barabási–Albert algorithm for generating random scale-free network using a preferential attachment mechanism (nodes with higher degree have stronger ability to grab a connection). Examples of generated scale-free graphs are given on figure 4.c. The plot of the distribution of the number of connections of our generated graph is given on figure 10.a.

The simulation showed that the scale-free banking system is prone to stronger contagion effect and the number of defaults rises sharply due to the presence of hubs (figure 11). It is important to note that for the scale-free system, while increasing the interconnectedness of the system, the defaults are dropping rapidly in comparison with the random network. .

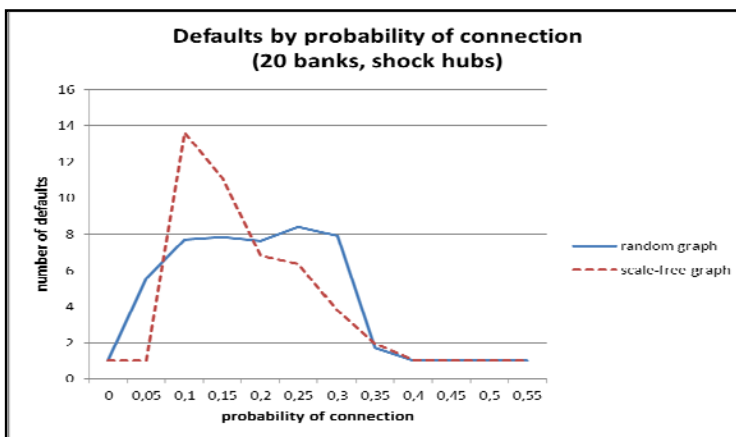


Figure 11. Behavior of the scale-free network (---) and the random (\_\_\_)



The assumption that the system has a scale-free underlying graph allowed the construction of a fundamentally new type of model for simulating the measures ensuring stability and efficiency of the banking market

For applying such model there could be implemented and refined few basic principles (Lewis, 2006) with the following interpretation:

*Principle 1:* You need a network to fight a network. This principle applied to the banking system may have the following two meanings:

(1) Systemic instability in the banking market has a network nature, so the means to oppose it should have a network character (a network of national supervisors - ESFS).

(2) Due to the size of European and international banking market only network approach would be effective. It is not economically feasible to protect every link in the system. European Commission studies show that the national Deposit Guarantee Schemes in the EU would not withstand the shock if several hubs (large, systemically important banks) fail.

*Principle 2:* Protect the hubs, not the connections. This principle is directly related to the preceding because the banking market is built on a network basis. Hubs are critical points, therefore they must be protected. Considering the scarcity of resources and the fact that such a network could be enormous in size, it is not possible to protect each node of the system, so efforts should be focused on the critical points.

*Principle 3:* Invest 80/20. The capital in the banking system is not equally allocated. One could say that the majority of the assets in the banking system are held by a small number of banks. This suggests that the 80/20 rule could be applied, i.e. 80% of resources should be invested in 20% of the units (which are critical to the system).

*Principle 4:* Asymmetric thinking. Innovation in the financial sector often is used as a way to avoid certain regulations. In this regard, regulatory approaches must evolve, and adapt to the market situation, and anticipate and manage the development of the system.

*Principle 5:* Dual purpose solutions. The scarcity of government resources raises the question for stakeholders' involvement in the process of seeking solutions for improving the stability of the banking system. Example, the establishment of joint entities for electronic or cash payments, creation of shock absorber funds and other initiatives with purely private capital, would increase the stability of the system, but would also help to improve its efficiency. Thus resources optimization could be achieved.

Taking into account these principles, we introduced in the simulation model **three protection strategies**:

*The first* consists of proportional allocation of bail-out funds to the banks. This strategy resembles to a theoretical government bail-out program where the banks' capital is increased. The result of its simulation is on figure 12.b.

*The second* one is a derivate of the first, but the bail-out funds are allocated only the biggest banks (the hubs). The result of its simulation is on figure 12.c.

*The third* strategy is called "toxic bank" and resembles to a theoretical government bail-out program, where there is a special institution buying the troubled assets from the banks while applying certain discount ratio. For better comparison we are using identical budget amount for all strategies. The result of its simulation is on figure 12.d.

We can see that the different strategies affect the contagion profile in a different aspect. The "proportional allocation" strategy reduces significantly the number of defaulted banks when the shock to the initial bank is moderate. The "hubs allocation" strategy gives overall lower number of defaulted banks. Nevertheless, this strategy is more effective when we have a full-scale shock. The "toxic bank" strategy turns out to be the most effective. It gives the lowest number of defaulted banks, independently of the shock size, because this strategy is the most flexible – the funds are allocated on a case-by-case basis covering only the troubled banks. Depending on the regulators or government's policy and the budget limits, a different discount ratio could be applied while buying toxic assets from the banks. Irrespective to their effectiveness and characteristics, all the strategies are bound with the same budget limits and when the shock gets enough big they cannot save system entirely. If the

funds limit is not sufficient to cope with the shock scale, these strategies can only postpone the system breakdown and give enough time for the economists and politicians to engage in more serious reforms.

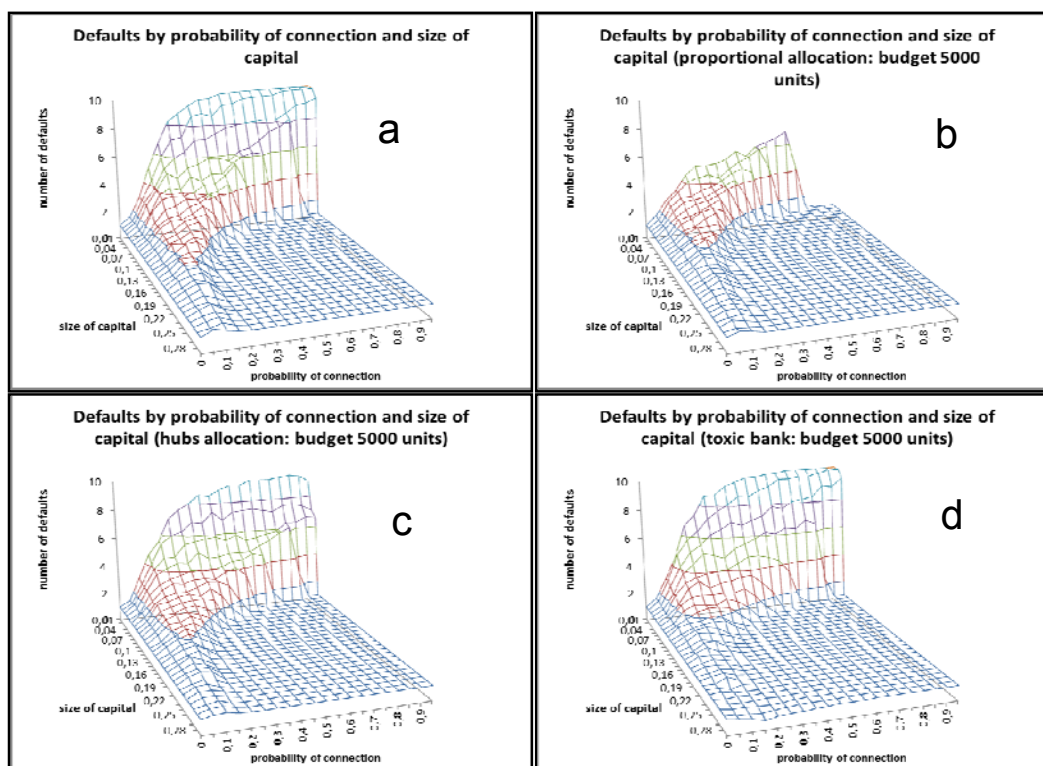


Figure 12. The non-protected system (a) and the system, protected with different strategies (b-d)

Looking at the overall contagion profiles of the different strategies, we can notice that the “proportional allocation” strategy is more effective when the system is significantly undercapitalized. This is due to the fact that the strategy increases directly the capital base of the banks in the system. The “toxic bank” strategy is more effective with moderately capitalized banks because it reduces the toxic portfolios without affecting their capital.

## CONCLUSIONS

The results obtained show that the implementation of the network approach to the banking system offers interesting opportunities for reorganizing its structure and predicting its response in crisis situations. This would contribute to the financial system transformation recommended in 2009 by the European Commission<sup>1</sup>, and for the establishment of a new system for European financial regulation and strengthening the cooperation and coordination between national supervisors.

The adoption of the banking system as a high-level economic infrastructure (separate critical infrastructure) allows the development of fundamentally new type of methods for ensuring stability and efficiency.

<sup>1</sup> European Commission, (2009), The de Larosière Group – Report, The High-Level Group on Supervision

---

## Bibliography

---

- Allen F., Babus A. (2008), Networks in Finance , University of Pennsylvania, 2008
- Elaterids, Hristian (2010), B. Sc. Diploma work, New Bulgarian University, Department of Computer Science
- Lewis, T., (2006), Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation, Wiley
- Mirchev L, Slavova V., Elefteridis H. (2010), Financial System Transformation - a network approach, in: proc. of 27èmes Journées d'Économie Monétaire et Bancaire • Groupement de Recherche Européen « Monnaie Banque Finance » Université Montesquieu - Bordeaux IV", Bordeaux, France, in print, online available on [http://gdre\\_mbf\\_2010.u-bordeaux4.fr/sites/gdre\\_mbf\\_2010/IMG/pdf/mirchev\\_slavova\\_elefteridis.pdf](http://gdre_mbf_2010.u-bordeaux4.fr/sites/gdre_mbf_2010/IMG/pdf/mirchev_slavova_elefteridis.pdf)
- Nier, E., Yang, J., Yorulmazer, T., Alentorn, A., Network Models and Financial Stability, Bank of England, 2008

---

## Authors' Information

---



**Velina Slavova**, professor, New Bulgarian University, Department of Computer Science, [vslavova@nbu.bg](mailto:vslavova@nbu.bg) (She coordinated the work and gave instructions. Also gave some algorithmic ideas and explained the results from a formal point of view. Hates programming, but finally implemented "Wave" just to see how to improve its complexity)



**Lyubomir Mirchev**, Ph. D. student, University of Nice – Sophia Antipolis, France, University of National and World Economy – Sofia, working at the National Bank of Bulgaria; [mirchev@gredeg.cnrs.fr](mailto:mirchev@gredeg.cnrs.fr) (The main figure. That's his idea, a part of his Ph. D Thesis in finance. Important: he is not bad in programming. He invented the protecting strategies.)



**Hristian Elefteridis**, B. Sc. Student, New Bulgarian University, Department of Computer Science, [hristian4o@nbu.bg](mailto:hristian4o@nbu.bg) (He realized almost all the programs within his diploma work and understood everything in finance. He is too familiar with all these new technologies...)

## BUSINESS INFORMATICS STUDY IN CEE COUNTRIES

Vasyl Grygorkiv, Ruslan Biloskursky, Andrii Verstiak, Igor Vinnychuk

**Abstract:** *The conditions of the IT-industry, opportunities and problems of preparation of IT-specialists in CEE-countries has been described. The ways of improvement of cooperation between IT-companies and universities for increases of IT-education quality and its level have been offered.*

**Keywords:** *business informatics, bologna study architecture, IT, IT-specialist, CEE-countries.*

---

### Introduction

---

Nowadays when Information and Communication Technologies (ICT) play an important strategic role for business success, enterprises are challenged to efficiently leverage their most valuable and underleveraged resource: the intellectual capital of their highly educated, skilled and experienced employees. This challenges academia to provide students with excellent education comprised of interdisciplinary Business Informatics competences and up-to-date research results.

The Bologna Study Architecture challenges universities and other higher-education institutions to change the structures of educational programs, to provide new academic degrees, to establish, adapt and enhance new quality frameworks and criteria and to increase the level of international cooperation both in research and in academic programs. The provision of new modular curricula in Business Informatics needs to tackle, especially at a Masters' Degree level, the topics of knowledge communication and evaluation, methods and content delivery in the context of technology-based knowledge transfer, quality criteria and performance indicators for education.

A raising number of cooperation between academia and industry on a research level provides both sides with beneficial, application-driven research results. The selection of appropriate research methods and tools is of utmost importance in this context especially in the view of industry-driven research initiatives as well as in empirical research. One aim of such concepts is to investigate approaches and offer for the storage and delivery of knowledge in enterprises that become essential pillars in the context of worldwide competition. Tightly linked to it is also the question of distribution and communication of research results as well as general legal conditions of intellectual property and interdependencies of law and technology.

Considering the impact on industrial projects, staffing requires fitting skill profiles and clear qualification requirements for enterprises. Therefore the business strategies in CEE countries and the alignment to the corresponding HR and IS strategies must be clearly defined. Nevertheless, cultural, social, economical and educational differences can be an additional hurdle for enterprises when they consider concepts like outsourcing and offshoring. From this point of view it is seen as especially interesting to also have Best Practices and Case Studies for ICT-applications in CEE countries.

---

### Business Informatics in Romania

---

Education in the field of informatics in Romania has a pyramidal structure, with specialized informatics and general education high schools at the base, BA and MA studies at the middle level and PhD studies at the top.

Historically, the first ones were the BA studies within the Departments of Computer Machinery at the University of Bucharest and BBU, at the Faculty of Mathematics in 1959. The computer science studies started in 1964 at the Faculty of Automation and Computers within the UPT and in 1967 at the PUB. Today, many state and private universities have Informatics and/or Computer Science Faculties.

Representatives of the openness in 1965, Grigore Moisil and some other intellectuals explained to Ceausescu that the economy could not work without informatics. In this respect, in 1967 CEPECA (the Centre for the improvement of the staff) was founded, financed by the UN that had the mission to prepare top managers in

Romania. The lectures were held by UN experts. These delivered lectures to the top managers to make them sensitive to computers, information systems and economic informatics (EI).

Some presentations were delivered concerning informatics and its usages between 1965 and 1969 at the Cybernetic Committee of the Romanian Academy, led by Grigore Moisil. The participants analyzed the curricula of the new Informatics Departments. The conclusion was that the curricula contained much more unusable mathematical theories and lacked in economic disciplines. For this 1967.

To illustrate the difference between informatics and EI they explain this to their students: "An e-shop can be made by students from computer science, informatics, you or even pupils from high school. You have to assure the knowledge about management, marketing, finance, accounting, simulation and forecasting, business plan etc. You need to be able to make an e-business by this."

The curriculum is conceived by respecting some basic principles as the following:

1. The preparation needs to be interdisciplinary, it covers 4 areas: economic sciences and business, computer science, intermediary disciplines, and complementary disciplines. 30-40 percent of our graduates are working in informatics, 20-30 percent are working as employees in different economic fields, and the rest have their own business in different fields.
2. The preparation needs to be differentiated for each level of study.
3. The preparation needs to assure a balance between theory and practice.

The BA prepares operational personnel and first-level managers. In Romania, the economic and business software companies are well developed and need specialized testers, programmers, technical assistants, data base and data warehouse administrators. Some of them have their own SMEs.

The MA prepares people for EI analysts, middle or top managers in the same or different economic and business fields. One of the alumni, who is an economic top manager of a German-Romanian company said: "I have to admit that since my graduation I have never written a program line, but EI has taught me to analyze differently the problems, data and knowledge in the company."

The PhD level usually prepares top managers, personnel for higher education, top analysts and EI developers, as ERPs, core banking systems, e-Business systems, economic and informatics advisor etc.

Some of our alumni are working abroad, from Australia to Canada. There are 30-40 of our alumni working only in Paris alone.

Referring to the curriculum, we can mention that:

The BA has 220 ETCS, 20 for the Diploma. 3 semesters assure the general preparation and the other 3, the specialization. The first 3 semesters represent 30% of the courses mainly in economic and business subjects (economics, management, marketing, accounting, and finances, international and European business etc.). Other 40% include informatics (algorithms and data structures, programming languages and environments, design of IS, 2 Data Bases courses, introduction to Artificial Intelligence, ERP/EAS, web design and Internet programming), intermediates 20 % (e-Business, testing of the programs and systems, 2 practical stages), 10% complementary (modern languages: main language English and secondary languages that can be German, French, Spanish, or Italian, business ethics, philosophy and sociology, history of economy etc.). 80% of the courses are core, 20% being optional. The BA exam consists in a Diploma Paper.

At the graduate level there are two kinds of curricula.

The one-year graduate Economic Informatics and Information Society has 70 ETCS, 10 for dissertation. This program is 50% informatics (C Sharp, CORBA and Java, Programming on components, Special chapters in Data Bases), 50% intermediates (special problems of e-Business, Decision Support Systems, Banking and finance information systems and EI systems engineering and reengineering).

A two-year master program totals 130 ECTSes, of which 10s are for the dissertation. The curricula contains: the first semester - economic core 20% (collaborative management, strategic marketing, finance administration), 10% - informatics; the second semester - 30% core informatics and intermediates (Knowledge Management,

Advanced programming environment, EI project management and design, Advanced Data Bases, Programming engineering and Artificial intelligence). The third semester is dedicated to the specialization and represents 30% of the curricula. The e-Business program includes e-Business design and development, m-Business, Grids and distributed systems, Virtual enterprises and business; the Decision Systems Program includes DSS design and development, Special issues in ERP/EAI, Simulation, Component programming and development. The last semester represents 20% and contains the Praxis and two specific optional courses. The graduate level is concluded by a dissertation. An interesting study was made in [1]. The author made a survey modeling the EI students profile based on questionnaires and data-mining.

The PhD School contains 5 courses. Two of them are from the EI: DSS and collaborative systems and Techniques to design and security of the EI systems; the other two contain economic or business subjects chosen by the students and their advisors. A general subject that is included in all the doctoral programs is Research Methodologies.

---

### **Implementing Bologna Architecture – The Case of Business Informatics Curricula in Poland**

---

In a Polish tertiary education system, study majors are a subject of standardization by the Ministry of Science and Higher Education. Universities have limited freedom of creating their own fields of study and their proposals have to be acknowledged by the Ministry.

The field of Business Informatics was never granted an independent formal recognition in the Polish tertiary education. It was always combined with another field of studies. Until nineteen nineties, the field of study concerning the topics related to Business Informatics was called Data Processing and Accounting. After the fall of communism and political and economic changes, also the educational system was also reorganized. New majors were developed including Informatics and Econometrics. This major is offered by faculties of economy or management in general universities, four specialized universities of economics, and various other higher education institutions.

The major Informatics and Econometrics covers wide range of topics related to the fields of Econometrics, Statistics and Informatics, from Mathematical Economics to E-Business and Computer Programming. Various efforts were undertaken in order to establish Business Informatics as an independent major field of study. So far they are unsuccessful. Linking up these fields of study is often blamed for an unsatisfactory number of students enrolling in this major. Some students' comments suggest that they are concerned about difficult courses in the wide range of topics. Particularly, students interested in a career in software development or IT consulting question the necessity to take courses on advanced mathematical methods. On the other hand, it is questionable that a future financial analyst has to study Software Engineering.

The form of unified first and second degree studies lasting 5-6 years was dominant in Polish tertiary education. Separate undergraduate and postgraduate studies were popular only in private education sector and in the form of extramural studies in state owned universities. However, these kind of studies were never considered prestigious in comparison with unified first and second degree studies.

In 2007, in Polish tertiary education, the system of unified first and second degree studies was abandoned following the adoption of the Bologna declaration. New teaching standards for separate undergraduate and postgraduate studies were published by the Ministry of Science and Higher Education, and tertiary education institutions were obliged to adjust their curricula to ministerial standards.

The implementation of the new standards led to the disintegration of well established curricula for unified studies. This process is widely discussed in Poland. Some academics claim that it will lead the deterioration of quality of the studies particularly in the field of hard science. In this case gaps in knowledge and skills from earlier stages of education will be very difficult to catch up. The consequence of this may be a reduction in the quality of education, especially at the master degree studies, due to the enrolment of students previously awarded a bachelor degree

in another unrelated field of study. However, the purpose of this article is not the criticism of the two level studies, but the discussion of its implications for teaching Business Informatics.

Additional problems were caused by the very late publication of the Decree of the Minister of Science and Higher Education including teaching standards for respective studies [2]. It was published in July 2007, whereas it was put into force already on October 2007, forcing the universities to develop their curricula without full knowledge of ministerial standards.

The Decree of the Minister of Science and Higher Education from July 12th 2007 [2] defines new learning standards for the first and second degree studies in 118 majors, including Informatics and Econometrics. The document does not include the detailed curricula but so called "learning content". However in many fields, including Informatics and Econometrics, the elements of learning content are directly reflecting the names of courses listed in the previous ministerial curricula for unified master degree studies [3]. The learning content specific for the area of Informatics and Econometrics in the undergraduate studies includes ten elements [2]: Mathematical Analysis, Linear Algebra, Descriptive and Economic Statistics, Probability Theory and Mathematical Statistics, Econometrics, Financial and Insurance Mathematics, Operational Research, *Computer Programming, Systems Design, Databases*.

Only three often topics are directly in the field of Business Informatics. Similar situation occurs in the case of postgraduate studies, only tree of nine learning contents are within field of Business Informatics: Multicriteria Analysis, Representative Sampling, Dynamic and Financial Econometrics, Mathematical Economics, Actuarial Methods, Statistical Analysis in Market Research, *Computer Networks, Software Engineering, Management Information Systems*.

A characteristic of this standard is little repetition of the learning content on both levels of education. As a result, students knowledge can be considerably different depending on which level of studies they carried out. Moreover, it is hard to find explanation why, for example, Computer Networks are taught at the second degree, whereas Computer Programming at the first degree. Such way of developing the program of study indicates incomplete understanding of the concept of separate undergraduate and postgraduate studies, allowing students to get both degrees in different fields. In fact, paragraph 12 of the very same document states as follows: "The completion of the first degree studies enables to continue education on the second degree studies in the same or different field" [2]. It can be assumed that the core courses from the curriculum of the unified first and second degree studies were split into learning contents of separate levels, without thorough consideration. The reason for this can be lack of understanding the idea of Bologna architecture, skepticism about the quality of two level studies or, simply, attachment to the traditional curriculum and unwillingness to change.

---

### **The Conditions and Perspectives of the IT Study in Ukraine**

---

At the moment Ukraine is facing the tasks of increasing the competitiveness of its economy, changing the structure of its gross domestic product (GDP) in favor of high-technology and science intensive sectors. It is impossible to solve these tasks without onrush of Information Technology (IT) sector due to the fact that it is basic for any economic sector. In the new economy the most valuable characteristics of workers are: informational competence: search and processing of information; creative and analytic way of thinking; ability to work with cycles and projects; ability to solve problems, personal responsibility; effective communication, team work; high efficiency.

The 1/8 of the budget of Ukraine goes to science and education. One can get higher education degree in 348 Universities, where 620 000 students study, this number also includes 30 000 students per year who graduate technical sciences.

Every year in Computer Sciences 1 000 people get a PhD or Associate Professor degree, 9 000 Master of Science, 4 000 - Bachelor of Science. Main centers of IT education in Ukraine are Kyiv (25000 students, 87 universities), Kharkiv (20000 students, 61 universities), L'viv (15000 students, 42 universities) [4].

Every sphere has the problem of training highly qualified professionals, but the IT sphere has the most significant problem [5].

This problem comprises of several factors: novelty and onrush of the sphere, variety of qualification requirements for professionals, high financial expenses on workshops equipment, difficulty organizing professional practice for students, and many others.

The IT evolution is faster than educational system can get used to the changes. New approaches and technologies arise every 10 month, fundamentally new concepts - every 3 years, change operational systems and hardware platform - every 5 years.

Universities do not have enough time to follow the development of IT sphere and, though, I professionals who graduate from the universities cannot go to work without extra skill Nevertheless, there is a lack of such professionals.

The companies are obliged to train their workers which require high financial expenses. There i high demand on financial, human resources and time which are spent on training professionals during their work. Not only IT companies suffer from the lack of qualified professionals, but also the consumers - all spheres of economy, state and population.

One of the biggest problems is also the situation with the professors' qualification degree. These problems can be solved by means of organization of mutually beneficial cooperation between universities and IT companies, as well as the development of research and educational resources.

It is essential for universities to cooperate with IT companies during the process of preparing professionals in order to define requirements that face graduates, to have an advice during drafting teaching plans, which correspond to the needs of modern IT industry, organization of students' practices etc.

---

## Conclusion

We consider that EI is very necessary in the future because of the globalization, the Internet and the Information and Knowledge Society, respectively, where the EI systems need well interdisciplinary prepared specialists. EI education is important for CEE-countries because, in our opinion, the large number of students specializing in finance, banking and other economic departments will decrease due to the current economic crisis. We estimate that the informatics activity will also be diminished because of the increase in salaries in IT in CEE-countries. Finally, this specialization offers our EI students diverse opportunities in the areas of informatics, economics and business.

---

## Bibliography

1. V.P. Bresfelean The IT implications about the academic institutions management, Risoprint, 2008.
2. Decree of the Minister of Science and Higher Education from July 12th 2007 on designation of teaching standards for respective studies and levels of education, as well as a procedure of creating and conditions which should be satisfy in order to run cross-field and macro-field studfish (in Polish), Dziennik Ustaw, 164,1166, 2007.
3. Decree of the Minister of National Education and Sport from April 18th 2002 on designation of teaching standards for respective studies and levels of education (in Polish), Dziennik Ustaw, 116,1004, 2002.
4. Ministry of Education and Science and materials from "2nd Forum on IT Education in Ukraine"
5. <http://www97.intel.com/odyssey/index.asp>



---

**Authors' Information**

---



**Vasyl Grygorkiv** – Economic and mathematical modeling department, chief, Chernivtsi National University, Kotsubinskiy st. 2, Chernivtsi, Ukraine; e-mail: [emmvsh@gmail.com](mailto:emmvsh@gmail.com)  
Major Fields of Scientific Research: Business Informatics, IT-technologies



**Ruslan Biloskurskyy** – Economic and mathematical modeling department, PhD Chernivtsi National University, Kotsubinskiy st. 2, Chernivtsi, Ukraine; e-mail: [billr32@mail.ru](mailto:billr32@mail.ru)  
Major Fields of Scientific Research: Business Informatics, IT-technologies



**Andrii Verstiak** – Economic and mathematical modeling department, PhD, Chernivtsi National University, Kotsubinskiy st. 2, Chernivtsi, Ukraine; e-mail: [averstyak@gmail.com](mailto:averstyak@gmail.com)  
Major Fields of Scientific Research: Business Informatics, IT-technologies



**Igor Vinnychuk** – Economic and mathematical modeling department, Research Assistant, Chernivtsi National University, Kotsubinskiy st. 2, Chernivtsi, Ukraine; e-mail: [igor.kapitoshka@gmail.com](mailto:igor.kapitoshka@gmail.com)  
Major Fields of Scientific Research: Business Informatics, IT-technologies

---

---

## Modern (e-) Learning

---

---

### СОВРЕМЕННЫЕ ПРОБЛЕМЫ ОБРАЗОВАНИЯ

Алексей Ф. Волошин

**Abstract:** *Рассматриваются современные проблемы образования, обсуждается концепция университета мирового класса. Анализируются проблемы образования в Украине, взаимодействия образования и воспитания, классического и e-образования, развития личности в условиях электронного рынка знаний. Приводится пример коллективной студенческой разработки учебно-методической, обучающе-тестирующей, оценивающей знания пользователя, системы.*

**Keywords:** *образование, e-образование, университет мирового класса, электронный рынок знаний.*

**ACM Classification Keywords:** *D.2.13 Reusable Software- Reuse models. K.3.1. Computer Uses in Education, Computer assisted instruction (CA).*

---

#### Введение

При подготовке конференции MeL-2010 автор беседовал с различными субъектами системы образования – студентами, аспирантами, коллегами (естественнонаучных и гуманитарных специальностей). Особенно запомнился один тезис при общении со студентами: «В первую очередь мы должны заботиться не об овладении предметом, а о сдаче экзамена, причем оценка в значительной степени зависит не от «объективных показателей», а от личностных качеств преподавателя». Где-то подобная (дословно) мысль мне встречалась... Вспомнил, нашел, цитирую: «Учащиеся должны заботиться не о том, чтобы получить образование, а о том, чтобы выдержать экзамен, причем он должен выучить заранее не только ответы на каждый вопрос, но и изучить пристрастия господ преподавателей, их настроение, что они предпочитают в науке и к чему питают отвращение». Приведенные слова принадлежат знаменитому французскому математику Эваристу Галуа [1811-1832] (см. [Дальма, 1984]). Возникает естественный вопрос – произошли ли за последние 2 столетия в системе образования существенные (сущностные) изменения? Безусловно, образование стало в значительной степени массовым, во многих странах начальное и среднее образование стали обязательными, в развитых странах существуют тенденции ко всеобщему высшему образованию. Однако, взаимодействие системы «ученик-учитель» не претерпело существенных изменений. Из слов лауреата Нобелевской премии Дж. Сакса, «до начала научной революции XVII ст. почти все народы существовали на грани выживания, однако после трех столетий технологического и научного развития выживание перестало быть правилом». Теперь время и силы человечества могут (и должны) быть направлены на реализацию предназначения *homo sapiens* – познания и творения мира и себя, к творчеству [Бердяев, 2006]. Если в индустриальную эпоху развитие общества происходило благодаря его отдельным представителям («талантам и гениям»), то в современную эпоху информатизации и глобализации каждый человек (свободный от борьбы за выживание, тяжелого физического и рутинного умственного труда) должен (и может) быть «творцом». В основе воспитания творческого человека должна лежать «творческая» система образования.

Поскольку автор доклада имел непосредственное отношение к формированию тематики конференции, его доклад будет представлен в виде комментариев к основным пунктам программы с попыткой наметить контуры «творческой» системы образования.

---

### Философия и методология современного образования

---

20-30 июня с.г. в Болгарии состоялся Форум конференций «Информационные теории и их приложения», на котором, в частности, была проведена дискуссия «Глобальные проблемы сообщества *homo sapiens* в современном информационном *flat world*». Основой дискуссии послужила книга-бестселлер Томаса Фридмана «The World is Flat» («Мир – плоский», [Фридман, 2008]). Одной из основных глобальных проблем современного мира называется проблема образования: «Единственный способ обеспечить рост благосостояния – построить общество, которое будет воспитывать людей, способных творить будущее. Для этого необходимо обеспечить «правильные» образование, инфраструктуру, стремления, руководство, воспитание» [Фридман, 2008, стр. 407]. Что понимается под «правильностью»? См. книгу. В докладе автор будет приводить некоторые определения «правильного образования» со своими комментариями и «проекциями» на систему образования в Украине, в частности, в Киевском национальном университете имени Тараса Шевченко, в частности, на факультете кибернетики, где работает автор. Будут приведены также некоторые предложения автора руководству факультета и университета, высказанные в последние 10 лет. Отдельные темы подробнее будут представлены в тезисах MeL-2010.

Оценка качества системы образования в Украине в многочисленных публикациях, выступлениях, обсуждениях, дискуссиях в последнее время оценивается от «больна» до «смертельно больна» (см. , например, «Зеркало недели» от 12.09.2009 – «Высшее образование или лохотрон»; см. многочисленные публикации 15-25 июля 2010 г. об организации вступительной компании в вузы Украины).

Идентифицируя свое мировоззрение с «мягким детерминизмом» [Фромм, 2009], автор придерживается точки зрения, согласно которой человек детерминирован законами причины и следствия, однако он может (и должен) создавать «сферу свободы» с помощью сознания, активной воли и готовности к борьбе. В частности, активная воля учителя (педагога, преподавателя, воспитателя) должна проявляться, в первую очередь, в пробуждении творческого сознания у ученика, сеянии доброго, истинного, вечного в обществе, борьбе с косностью, глупостью, подлостью, злом, «невзирая на лица».

Поэтому, оценивая уровень образования в Украине на «три с минусом», автор считает, что сознание, воля и борьба должны привести к успеху, хотя бы «локальному».

---

### World-Class University

---

В декабре 2009 года в Шанхае (КНР) состоялась третья международная конференция по вопросам создания «университета мирового класса» (университета XXI ст. , университета будущего). Какого-либо заметного отклика в Украине она не вызвала, во всяком случае ни один из двух десятков коллег автора о проекте *World-Class University* (WCU) и не слышал. Автор ознакомился с материалами конференции, резюме – Украина может выпасть из мирового процесса модернизации образования. WCU определяется тремя группами качеств: 1) Высоким уровнем интеллекта преподавателей и студентов; 2) Значительными финансовыми возможностями; 3) Оптимальной управленческой моделью.

Первое и главнейшее умение, которое необходимо развивать в WCU – это умение «учиться учиться». Важно не только чему вы учите или что вы знаете, но и как вы учите и как вы учитесь. Автор доклада в своей педагогической практике старается придерживаться принципа выдающегося ученого и педагога П.Л. Капицы – «Лекция должна быть интересной, понятной и...». Если два первые критерия студенты и преподаватели, с которыми автор общался, называли со 2-3 раза, то ни один из десятков опрошенных коллег третий критерий не назвал и с десятого раза, а студенты называли его с 4-5 раза. Этот критерий – «веселой». В [Фридман, 2008] приводится формула успешного обучения: заинтересованность плюс увлечение важнее уровня интеллекта.

Автор полностью согласен с мыслью, что в XXI ст. роль классического образования будет только возрастать, у творческого человека в равной степени должны быть развиты оба полушария. Вице-министр образования Китая [Фридман, 2008]: «Для того, чтобы люди умели творить и самостоятельно мыслить необходимо единство гуманитарного и естественного образования»; ректор Стенфордского

университета: «Кеннеди хотел отправить человека на Луну. Я мечтаю отправить каждого американца в университет»; Рич Де Мило (бывший директор фирмы Hewlett-Packard, нынешний декан Технологического университета Джорджии: «Сама идея отдельных факультетов является безумной» (это, наверное, все же через чур!). Особенно следует отметить, что необходимость людей с базовой математической подготовкой будет возрастать, Ассоциация «Американский диплом»: «Следующие 10 лет в Америке в 62% сферах деятельности условием приема на работу будет знание математики».

За 20 лет независимости существенных изменений в организации высшего образования в Украине не произошло. Введение контрактной формы обучения привело лишь к снижению образовательного уровня выпускников (контрактников практически невозможно отчислить за неуспеваемость). Но и обучающихся по государственному заказу тоже отчислить (за неуспеваемость, прогулы) непросто – государственный заказ необходимо выполнять.

Предложения (из архива автора): 1) На первый курс зачисляется на 15-20 % больше планового числа абитуриентов («стажеров») без предоставления стипендий и общежитий, после 1-го курса «лишние» отчисляются. Многолетний педагогический опыт автора, да и мировая практика, свидетельствуют, что именно столько лиц по тем или иным причинам не могут овладеть программой; 2) Требуется пересмотра и система платного обучения. Заслуживает внимания система государственного кредитования – выпускник обязан выплатить кредит за 5-10 лет после окончания вуза, при отличной успеваемости кредит частично (или даже полностью) погашается.

Значительным источником финансирования в вузах США является прием иностранных студентов. Так, по данным Института международного образования США в 2004-2005 гг. в университеты США поступило более 80 тыс. граждан Индии, 60 тыс. Китая, 50 тыс. Южной Кореи. В начале последнего десятилетия на факультет кибернетики КНУ поступило несколько десятков китайских студентов. Они были приняты без знания украинского языка (как впрочем, русского и английского) и математики. Автор официально предупредил руководство университета, что делать этого нельзя, поскольку невозможно в этих условиях подготовить более-менее полноценного специалиста, что станет ясно и китайской стороне. Это и случилось – китайские студенты на факультет больше не поступают. Резюме (из архива автора, 2009 г. – 1) Не допускать преступной практики бывшего руководства КНУ по зачислению на естественнонаучные факультеты без знания языка и предмета; 2) Стимулировать естественнонаучные факультеты к переходу на преподавание иностранным студентам на английском языке.

И, безусловно, образование в университете должно быть неразрывно с научной деятельностью. Статус КНУ «научно-исследовательский» предполагает «обеспечение неразрывной связи образования, науки и инновационной деятельности» (ректор КНУ проф. Л. Губерский, «Зеркало недели» от 10.10.2009 г.). Причины разделения образования и науки в СССР еще более просты, чем приведенные проф. С.Квитом («ЗН» от 28.02.09 – «Контроль интеллектуалов и концентрация усилий на стратегически важных направлениях»). Это элементарный режим секретности, вспомним «акт на возможность публикации», «подписку о неразглашении» и т.п. Ведь 90 % научных исследований финансировались через ВПК (военно-промышленный комплекс).

Особенно следует остановиться на подготовке научных кадров «высшей квалификации» через аспирантуру и докторантуру – в вузе это было и остается основным «рычагом» развития науки. Но, как известно (см. статью С. Квита), сегодня меньше 30% всех аспирантов вообще когда-либо выходят на защиту, лишь 7% делают это в срок. Полностью поддерживая «стратегические» предложения С. Квита, приведу «тактические», не требующие кардинальной перестройки образования. Эти предложения автор неоднократно высказывал на ученых советах, руководству университета в течение последних десяти лет: 1) Практиковать продолжение срока аспирантуры на 1-2 года (за счет невыполнения плана по набору и отчислений). В 20-х годах в СССР, а сейчас и во всех цивилизованных странах отсутствуют «жесткие» сроки пребывания в аспирантуре (докторантуре). Даже, если работа и подготовлена в срок, защита затягивается минимум на полгода – в это время аспирант является или безработным, или, наоборот, загруженный работой. Одновременно должна быть запрещена постоянная работа аспиранта – в США,

например, аспирант с разрешения деканата может работать только в каникулярное время. Сейчас не единичные случаи, когда аспирантура рассматривается как дополнительная «подработка» и место в общежитии; 2) Стимулировать научных руководителей с «высокой эффективностью» - например, 25 часов из 50 за руководство аспирантом защищать как аудиторные. Сейчас профессор, у которого 5 аспирантов и тот, у которого годами их нет (по данным ректората КНУ таких 40%), имеет одинаковую аудиторную нагрузку; 3) Стимулировать научных руководителей, например, финансированием командировок на научные конференции и т.п.; 4) Не разрешать осуществлять научное руководство лицам с «низкой эффективностью», например, меньшей, чем 33%. На факультете кибернетики имеются научные руководители, которые за много лет не подготовили ни одного кандидата наук, тем не менее, каждый год они набирают аспирантов. При анализе работы руководителей необходимо учитывать общее количество аспирантов, поступивших в аспирантуру, закончивших ее и защитивших диссертацию. Если за время пребывания на контракте (5-7 лет) на должности профессора под его руководством не защищено ни одной диссертации, он переводится на должность доцента, ведь во всем мире основной обязанностью профессора является подготовка кадров высшей квалификации; 5) Рекомендовать на звание профессора при наличии минимум двух подготовленных кандидатов наук, учебного пособия с грифом Министерства образования и монографии. Интересно отметить, что подобные нормы действовали с начала 2009 г. до середины с.г., сейчас они отменены, цель отмены понятна – скоро все чиновники станут профессорами.

### **Международные рейтинги, Болонская система**

Сам термин *World-Class University (WCU)* предполагает наличие глобального пространства, универсальных связей университетов мира, согласования принципов содержания и формы учебного процесса и научной деятельности. Одним из основных инструментов интеграции образования являются международные рейтинги. С одной стороны, они служат для сравнительной оценки работы вузов, формируют рыночную среду образовательных учреждений, с другой стороны, показывают направление усовершенствования и реформирования работы, формируют модель «идеального» (*World-Class*) университета. История развития международных рейтингов коротка. Самыми известными являются «Шанхайский» (впервые опубликованный в 2003 г.) и «Таймс» (2004 г.). Судя по доступным автору публикациям, методы расчета рейтингов нуждаются в существенном усовершенствовании, существующие основаны на аддитивных свертках показателей (критериев) и носят в значительной степени эвристический характер. И главное – в системе критериев должны преобладать «результатирующие динамические критерии («по делам их узнаете их»), а не «потенциальные, статические» (количество академиков, докторов и т.д.). Так, публикуемый в последние годы рейтинг «200 лучших вузов Украины» («ЗН» от 30.05.09) состоит из трех критериев – индекс качества персонала (от 0 до 50%), индекса качества образования (0-30%), индекса международного признания (0-20%). В свою очередь, для формирования указанных индексов использовали 16 «индексов прямого назначения». Возникает масса вопросов- почему выбраны такие индексы, такие диапазоны их изменения и т.п.? Почему берется аддитивная свертка? Как неточность измерения индикаторов влияет на конечный рейтинг? И тому подобное. Между тем, построение коллективных порядков – это раздел теории принятия решений [Волошин, 2006], где получено ряд интересных научно-обоснованных результатов, ориентированных на решение задач большой размерности, с нечеткими данными и т.п. [Волошин, 2009,2010].

Решать конкретные проблемы интеграции в европейское образовательное пространство призвана «Болонская система». Автор считает «Болонскую систему» в целом прогрессивной, если ею руководствоваться не только «на бумаге». Главный элемент – кредитно-модульная система стимулирует работу студента на протяжении семестра. Студент, который не набрал за семестр 20 баллов (если на экзамен выносятся 40 баллов, удовлетворительная оценка – минимум 60 баллов), не должен допускаться до экзамена. К сожалению, у нас это не так – аргументацию руководства см. выше. И вообще, нужны четкие правила организации учебного процесса – если студент без уважительных причин пропустил определенное количество занятий, он отчисляется автоматически. Аналогично, если в сессию студент получил две неудовлетворительные оценки. Иначе, у молодежи воспитывается «правовой нигилизм»

(более того, цинизм) – одного можно наказать, другого за те же прегрешения – помиловать. Другой элемент «Болонской системы» - согласование учебных планов, объемов курсов, тоже, безусловно, является важным. Уменьшение в три раза количества специальностей по сравнению с европейскими вузами является объективным также.

---

### Образование и воспитание

---

Возвращаясь к «безумной» идее отдельных факультетов, отметим, что у нас она «вдвойне безумная». Научная кооперация между факультетами практически отсутствует. А сейчас, как никогда, необходимо объединение усилий математиков, кибернетиков, философов, психологов, филологов и т.д. для изучения системы «Человек» [Фромм, 2009]. Фрейдовский *homo sexualis* – изолированный самодостаточный человек, который вступает в отношения с другими ради удовлетворения собственных потребностей. Классический *homo economicus* имеет экономические нужды, которые находят удовлетворение в коллективной деятельности по обмену товарами на общественном рынке. Что есть *homo informaticus*? Человек, имеющий потребности в знаниях, которые он получает без прямого взаимодействия с другими участниками электронного рынка знаний? Человек виртуальный, растворившийся в коллективном (всемирном) интеллекте, потерявший свою индивидуальность? Кто прав – К. Маркс, который утверждал, что «не познание людей определяет их бытие, а напротив – это их общественное бытие определяет сознание»; или, напр., герой А. Турчинова [Турчинов, 2007]: «Сознание формирует окружающий нас мир»? Автор согласен с мыслью [Марков, 2009], что «электронный рынок знаний, вероятно, будет не только мешать развитию отдельных личностей, но и, возможно, окажет определенное отрицательное воздействие на развитие общества в целом». Более того, автору кажется, что в приведенной цитате следует убрать слова «вероятно» и «возможно». Поэтому автор поддерживает мысль ректора Калифорнийского технологического института Д. Балтимора: «Необходима настоящая революция в отношении родителей к образованию детей. Они должны быть готовы к «суровой» любви: настало время, когда необходимо оторвать ребенка от компьютера и телевизора и заставить его работать».

---

### Успехи наук и образование

---

Сейчас следует вернуться к вопросу, поставленному во введении, (что изменилось за последние 200 лет в образовании?). и посмотреть на него с другой стороны. Науке стало многое известно в строении, функционировании головного мозга человека (см., например, [Хокинс, 2007]), но как эти знания используются в системе образования (научения)? Ответ – практически никак. Ни один из десятка опрошенных мною коллег не слышал об «Инженерной педагогике», хотя в Европе с 1972 года существует Международное сообщество по инженерной педагогике. Книга президента сообщества А. Мелецинека (Австрия) имеет большой успех в Западной Европе [Мелецинек, 1998], в которой на основании концепции дидактики, методики, психологии и социологии даются практические рекомендации преподавателям – как готовиться к лекции, завоевывать аудиторию, качественно преподавать и т.д. В настоящий момент многие рекомендации несколько устарели, особенно, что касается технического обеспечения, но знания о скорости восприятия информации человеком (скорость «апперцепции») оперативной и долгосрочной памяти человека, емкости оперативной памяти, скорости забывания информации, зависимости запоминания от времени и повторений и т.д. и т.п. остаются неизменными. Этот пример иллюстрирует тот факт, что новые знания в практику образования поступают очень медленно. Более того, отсутствуют какие – либо стимулы овладения ними! Автор доклада был поражен безразличием абсолютного большинства коллег к проблемам современного образования. Оказывается все просто – пришел в аудиторию, что-то рассказал, принял экзамен, поставил оценку (самому себе). Правда сейчас (с введением Болонской системы) начали заставлять писать какие-то планы, приводить модульные контрольные и т.п. Автору знакомы педагоги, которым вообще безразлично, сколько на лекции студентов. Что здесь предпринять тактически? Нужен элементарный контроль – соответствие фактического материала на лекции рабочим

планам, посещаемости студентов, экзамен должна принимать комиссия (в идеале – проводить компьютерное тестирование, см. раздел «е-образование»).

---

### **е-образование**

В двух предыдущих разделах косвенно упоминалось об информационных (электронных, компьютерных) технологиях в образовании (е-образование). Существуют различные отношения к е-образованию, вплоть до противоположных. Выше автор упоминал о своем «мягком детерминизме», поэтому не отрицать, не считать его панацеей не намерен. Нужна взвешенная позиция, которая помогла бы разрешить противоречие «стандартизация обучения и индивидуальность обучаемых» [Марков, 2009]. Информационные технологии в образовании следует рассматривать как инструментарий интенсификации обучения, избавление его от технической, рутинной работы, как механизм коллективного обучения в процессе коллективной деятельности. Автор последние годы практикует коллективные задания при выполнении лабораторных работ по курсу «Теория принятия решений» с целью создания учебно-методических программных продуктов. Была разработана учебно-методическая, обучающе-тестирующая программная система [Волошин, 2008].

Другая основная функция информационных технологий в образовании – «объективизация» процесса оценивания знаний учащихся (см. Введение). В [Волошин, 2010] описан «сценарно-прецедентный» подход к оцениванию результатов. В тезисах конференции MeL-2010 будет описана окончательная версия программной системы как учебно-методической, обучающе-тестирующей и оценивающей знания.

В заключение – о взаимодействии классического и е-образования. Еще раз следует подчеркнуть – информационные технологии являются вспомогательными, инструментарием. Автор согласен с [Зайнутдинова, 2008], что основным звеном дидактического цикла в вузе является лекция, рождающаяся на глазах слушателя, в диалоге с ним. Непосредственное общение «ученик-учитель» можно дополнить, усовершенствовать, но не заменить. К. Маркс: «Нужда в другом человеке воспринимается как ощущение потребности в величайшем богатстве».

---

### **Заключение**

Развитие сферы образования в современном мире является одной из важнейших проблем человечества, разумное решение которой с помощью современных информационных технологий позволит человечеству выйти на новый уровень познания мира и себя, к созидательному творчеству. Для этого необходимы сознательная деятельность сообщества по формированию разумных стремлений, созданию оптимальной управленческой модели, построению соответствующей инфраструктуры и совершении «революции» в отношении родителей к образованию детей.

---

### **Благодарности**

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

---

### **Литература**

- [Дальма, 1984] Дальма А. Эварист Галуа, революционер и математик. Пер. с франц.-М.: Наука, 1984.-112 с.
- [Бердяев, 2006] Бердяев Н.О назначении человека.–М.: АСТ: АСТМОСКВА: ХРАНИТЕЛЬ, 2006.- 478 с.
- [Фридман, 2008] Фридман Т., Світ плаский! Глобалізований світ у XXI столітті. Пер. з англ. Київ: «АКТА», 2008.-676 с.
- [Фромм, 2009] Фромм Э. Кризис психоанализа.–М.: АСТ: АСТМОСКВА., 2009.- 252 с.
- [Турчинов, 2007] Турчинов А. Тайная вечеря.–К.: Криница, 2007.- 240 с.
- [Фромм, 2009] Фром Э. Душа человека.–М.: АСТ: АСТМОСКВА: ХРАНИТЕЛЬ, 2009.- 251 с.
- [Марков, 2009] Марков К., Личностное развитие в условиях электронного рынка знаний //Материалы международной научной конференции «Информационные технологии будущего и современное электронное обучение «Modern IT

- & (E-) Learning», 2009, Астрахань: ООО «Типография «НОВА», 2009.- 352 с.
- [Хокинс, 2007] Хокинс Дж. Об интеллекте: Пер. с англ.- М: ООО «И.Д. Вильямс», 2007.– 240 с.
- [Мелецинек, 1998] Мелецинек А. Инженерная педагогика, Пер. с нем.- Москва: МАДИ(ТУ), 1998.– 185 с.
- [Волошин, 2008] Волошин А. и др. Об опыте коллективной разработки учебно-методических программных систем // International Book Series "Information Science and Computing", N8, Sofia, Bulgaria, 2008. - P. 97-103.
- [Зайнутдинова, 2008] Зайнутдинова Л., Семенова Н. Интерактивная структура мультимедийной обучающей системы лекционного курса электротехнической дисциплины// International Book Series "Information Science and Computing", N8, Sofia, Bulgaria, 2008. - P. 111-116.
- [Волошин, 2010] Волошин А., Антосяк П. Процедура нахождения строгого результирующего отношения на множестве альтернатив в задаче линейного упорядочения альтернатив// Natural and Artificial Intelligence? ITHEA, Sofia, Bulgaria, 2010. - P. 78-84.
- [Voloshyn, 2009] Voloshyn O., Antosiak P. Procedures of Sequential Analysis and Sifting of Variants for the Linear Ordering Problem";.In: International Book Series "Information Science and Computing", 2009. -V. 3/2009, N15. - P. 149-154
- 

### Информация об авторе

---



**Волошин Алексей Федорович** –, профессор факультета кибернетики Киевского национального университета имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: [ovoloshin@unicyb.kiev.ua](mailto:ovoloshin@unicyb.kiev.ua)

**Сфера научных интересов:** принятие решений, системы поддержки принятия решений, экспертные системы, математическая экономика, е-образование



## ПРОБЛЕМЫ ВХОЖДЕНИЯ УКРАИНЫ В ЕВРОПЕЙСКОЕ ОБРАЗОВАТЕЛЬНОЕ ПРОСТРАНСТВО

Олег К. Закусило

**Аннотация:** В докладе дается обзор обстоятельств создания Европейского образовательного пространства и причин, препятствующих вхождению в него украинских университетов. Анализируется влияние основных участников на форму и содержание обучения в современной Украине. Проводится сравнение учебных планов украинских университетов с планами университетов других стран. Проводится анализ политики других стран по привлечению иностранных студентов. Оцениваются перспективы исследовательских университетов.

**Abstract:** The report reviews the circumstances of formation of the European higher education area as well as the factors retarding entry of Ukraine into this structure. It analyzes the influence of the main parties governing form and content of education under the conditions of the present-day Ukraine. A comparison of curriculums of Ukrainian universities with that of other countries is carried out. A detailed analysis of the policy pursued by other countries for the benefit of engaging students from abroad is given as well. The perspectives of research universities in Ukraine are discussed.

**Ключевые слова:** Европейское образовательное пространство; Болонский процесс; мобильность студентов; обучение иностранцев; проблемы вхождения в Европейское образовательное пространство; исследовательские университеты.

**Keywords:** European higher education area; student mobility; Bolonsky process; training foreign persons; problems of entering the European higher education area; research universities.

**ACM Classification Keywords:** H.5.3 Group and Organization Interfaces. K.3.1. Computer Uses in Education, Computer assisted instruction (CA).

---

### Введение

«Болонским» называют процесс создания единого европейского пространства высшего образования [1]. Он является проявлением интеграционных тенденций, наблюдающихся в этой части мира. Принято считать, что в истории Болонского процесса можно выделить три этапа: предыстория: от Великой хартии университетов (1088 г.) до Болонской декларации; начало: Болонская декларация (1999 г.); развитие: после Болонской декларации. По мнению специалистов, Болонский процесс является следствием как социальных, так и чисто академических и экономических проблем.

---

### Социальные обстоятельства возникновения Болонского процесса

В 20-ом столетии в ряде стран происходили студенческие протесты как по политическим мотивам, не имеющих непосредственного отношения именно к студентам, так и против законов или реформ образования. Сильнейшим из них было студенческое движение во Франции в мае 1968 года, когда страну захлестнули массовые беспорядки, использованные многочисленными группами (коммунистами, анархистами, правыми либертарными активистами) для реализации своих политических идей. Это движение перешло во всеобщую забастовку, на несколько недель парализовавшую всю страну. В мае-июне 1968 года во время столкновений с полицией было убито 4 манифестанта.

События в Париже вызвали волну студенческих протестов по всему миру. Западногерманское студенческое движение принимало участие в демонстрациях против «чрезвычайного законодательства». В Испании демонстрации против диктатуры Франко были разогнаны полицией. В Мехико студенческая

демонстрация закончилась расстрелом демонстрантов 02.10.1968 г. и стала известной как «резня в Тлателолко». Даже в Пакистане студенты вышли 7 ноября на улицы, протестуя против политики в сфере образования. Полиция открыла огонь и был застрелен учащийся колледжа.

Протесты и забастовки против законов или реформ образования во Франции имели регулярный характер. Самой продолжительной была трехмесячная забастовка в 1976 г., ставившая целью помешать проведению реформ второго цикла образования (создание новых специальностей). Это воспринималось студентами как попытка подогнать высшее образование под запросы рынка и сделать его еще менее доступным. Значительное влияние на это движение имела Революционная коммунистическая лига, в нем родились будущие политические кадры Организации коммунистов-интернационалистов, Лиги коммунистов-революционеров, Союза студентов-коммунистов. Когда же правительство Пьера Моруа в 1983 г. отменило ограничения при поступлении в вузы, уже правые и ультраправые студенты в середине мая организовали волнения в элитарных университетах, продолжавшиеся две недели. Через три года правительство Жака Ширака снова попыталось ввести ограничения при поступлении в вузы и сотни тысяч студентов вышли на улицы. Манифестации приобрели ожесточенный характер, погиб один студент-араб. Министр образования Ален Деваке был вынужден отступить и отменить реформу. Французские студенты снова поднялись в марте 1994 г. против «Контракта трудового содействия», уменьшавшего размер оплаты труда до 80% его начального значения для лиц до 25 лет, потом в ноябре-декабре против сокращения бюджетного финансирования высших учебных заведений [2].

---

### **Академические обстоятельства возникновения Болонского процесса**

---

Академическая предыстория Болонского процесса начинается с 1954 года, когда в Париже государства-члены Совета Европы подписали Европейскую культурную конвенцию, в которой подчеркивалась необходимость проведения политики совместных действий, направленных на сохранение европейской культуры и поощрение ее развития. Конвенция указывала на необходимость привлечения граждан всех государств к изучению языков, истории и культуры других стран и общей для них европейской культур. Самый старый в Европе Болонский университет не случайно стал инициатором межправительственной международной реформы в 1999 году. Еще за тринадцать лет до этого в 1986 году этот университет выступил с предложением создания Великой Хартии университетов (Magna Charta Universitatum). Хартия была разработана и окончательно подписана в Болонье в 1988 году во время торжественного празднования 900-летия Болонского университета. Следующим шагом стала Сорбонская декларация 1998 года, принятая при праздновании 800-летия парижского университета. Сам Болонский процесс обязан своим названием так называемой Болонской декларации, подписанной 19 июня 1999 года в этом итальянском городе министрами высшего образования 29 европейских стран. С этого момента начинается официальная история реформы, окончание которой зафиксировано на 2010 год.

Основные принципы Болонского процесса хорошо известны и даже беглый их анализ свидетельствует о том, что моделью будущей гармонизированной структуры европейского высшего образования служила типичная для США и Великобритании двухуровневая система (бакалаврат-магистратура), в разных вариантах применяющаяся в большинстве европейских, и не только европейских, стран. Для этого есть свои причины, которые, по нашему мнению, являются главными в реализации проекта под названием «Болонский процесс».

---

### **Экономические обстоятельства возникновения Болонского процесса**

---

В наше время термин «образовательное пространство» все чаще заменяется другим - «сфера оказания образовательных услуг». Это является следствием приобретения образованием признаков производства и наличия больших денежных масс, вращающихся в этой сфере. Для создания представления об этих суммах приведем пример Австралии – в бюджете этого государства доходы от привлечения к обучению

иностранных студентов занимают третье место. В США в 2007 году доходы от обучения иностранных студентов превышали доходы от продажи оружия. К моменту утверждения идей Болонского процесса уже пришло осознание существования мирового рынка оказания образовательных услуг и жестокой борьбы на нем. Пришло осознание и того, что на этом рынке университеты Западной Европы стали проигрывать университетам США. Определенным ответом на эти вызовы времени и стал Болонский процесс. О достигнутых результатах можно судить на примере Испании. В этой стране доля иностранных студентов возросла с 0,7% в 1995/96 учебном году до 1,7% в 2005/06. Особенно возросла доля иностранцев, обучающихся по постуниверситетским программам – с 5,2% до 22,1%. Поступлению иностранцев в испанские университеты способствует хорошо организованный порядок оформления, недорогое государственное образование, возможность получения стипендии, совмещение работы с учебой и комфортные условия проживания. Испанские университеты, как правило, резервируют 5% мест для студентов из стран, не входящих в Евросоюз.

За последние десять лет количество студентов, получающих образование вне страны проживания (мобильных студентов), возросло почти на 70% [3]. По данным Института статистики ЮНЕСКО, таких студентов в 1999 году насчитывалось 1,68 миллиона, а в 2008 году их оказалось более 2,9 миллиона. По прогнозам специалистов, в 2025 году их количество достигнет 2,7 миллиона. Большинство иностранных студентов обучаются в странах Европы и Северной Америки. Их доля от общего количества студентов составляет около 2% в Европе и около 3% в США. Но в некоторых университетах Великобритании, Франции и Германии количество иностранных студентов превышает 10%, а в отдельных университетах США достигает 30% от общего количества студентов.

В течение последних 30 лет основными поставщиками студентов на мировой рынок образовательных услуг являются страны Азии. Доля азиатских студентов, обучающихся за границей, составляет около 50% от общего количества иностранных студентов. Лидируют в этом списке Китай, Республика Корея, Малайзия и Индия. Наблюдается постоянный рост иностранных студентов из Таиланда, Индонезии, Сингапура и Пакистана. Вторым крупным донором иностранных студентов являются арабские страны - Марокко, Иран, Иордания, Алжир, Палестина.

Самыми популярными среди иностранных студентов, обучающихся, например, в США, являются программы по бизнесу и менеджменту (изучают 19,5% от общего количества иностранных студентов), техническим специальностям (18%), математике и вычислительной технике (9%), естественным наукам (8,5%), социальным наукам (7,5%), медицине и здравоохранению (4%). При этом много студентов отдают предпочтение получению базового образования в стране постоянного проживания, а продолжать обучение в магистратуре или аспирантуре за границей.

В последние годы серьезные меры по привлечению иностранных студентов предпринимаются правительствами Республики Корея [4], Новой Зеландии, Канады, КНР. В Новой Зеландии поступления в казну от этой отрасли занимают четвертое место. По данным Минобразования КНР, в этой стране в 2007 году количество иностранных студентов превысило 190 тысяч. Они приехали из 188 стран и регионов и учились в 544 китайских вузах. Правительство Великобритании планирует увеличить количество иностранных студентов в 2007-2012 годах с 300 тысяч на еще 100 тысяч. Эти студенты приносят ежегодно 4 миллиарда фунтов стерлингов вузам Великобритании и 10 миллиардов экономике страны в целом. Это превышает доходы от продуктов питания, табачной продукции, страхования, судостроительства и конструирования самолетов.

Данные о количестве студентов-иностранцев в 2007 году таковы (в тысячах лиц): США – 582,9, Великобритания – 376,2, Германия – 261, Франция – 247,5, Австралия – 202,4, Китай – 144, Япония – 120, Канада – больше 100, Сингапур – 86, Новая Зеландия – 67, Малайзия – 65, Республика Корея – 63,9.

Первый вице-премьер РФ Игорь Шувалов считает [5], что привлечение иностранных студентов должно стать одним из главных критериев оценивания работы высшего учебного заведения. По его мнению, Россия, доля которой в экспорте образовательных услуг не превышает 3% по количеству студентов, будет проигрывать глобальную конкуренцию, если не сделает решительные шаги по привлечению

талантливой молодежи. Он подчеркивает, что привлекательность для иностранцев не только повышает престиж вуза, но и помогает получить немалые инвестиции как самому университету, так и соответствующему региону. Вместе с тем, финансы не являются единственной целью интеграции в мировое образовательное пространство. Россия, как, кстати, и Украина, имеет серьезные демографические проблемы. Экспорт образовательных услуг в определенной степени решает геополитические задачи и наращивает научный капитал.

Нельзя сказать, что рассматриваемая тема остается совсем не замеченной высокими должностными лицами нашей страны. По подсчетам Министерства образования и науки Украины, наши вузы могут обучать втрое больше студентов, чем сейчас. «Многие страны за счет иностранцев формируют львиную долю бюджета страны. В первую очередь, это характерно для США, где учится 28% иностранных студентов. Ежегодно на иностранных студентах они зарабатывают 14 миллиардов долларов, а на экспорте оружия – 13,4 миллиарда» - считает начальник департамента Минобразования Е. Сухолиткий. По его данным, в 2007 году в Украине обучалось около 37 тысяч студентов из 129 стран, от их обучения Украина получила 89 миллионов долларов.

---

### **Обстоятельства присоединения Украины к Болонскому процессу**

---

Приходится констатировать, что вышеуказанные факторы, породившие Болонский процесс, не имеют отношения к Украине. У нас не было ни мощных студенческих движений, ни инициатив университетов, ни четко выраженных прагматических мотивов. Мы подписали документы о присоединении к процессу, формировавшегося без нашего участия. Присоединение к Болонскому процессу диктовалось политическими соображениями, связанными с романтическими представлениями о мгновенном вступлении в Европейский Союз. Учитывая украинскую ментальность, это имело следствием желание высоких должностных лиц доложить руководству страны, что какой-то частью мы уже начали входить в Европу и становимся все большими европейцами. Вместе с тем, в отличие от прагматичных европейцев, мы сохранили особенности национальных способов организации работы и не создали ни тщательно проработанных календарных планов, ни экономического обоснования этого грандиозного проекта, ни соответствующей нормативно-правовой базы. Полностью отсутствуют проекты по привлечению в Украину иностранных студентов. Отсутствует фактически все, что свидетельствовало бы о стремлении действовать в интересах нашей экономики и, в конечном счете, всего общества. Тезис о мобильности студентов извращен до неузнаваемости – оказывается, мы должны создавать условия для возможности обучения наших студентов за границей, а необходимость привлечения иностранцев к обучению у нас замалчивается. Абсолютно незаметны усилия, направленные на повышение качества нашего образования. Настоящая автономия университетов в нашей стране отсутствует, хотя это один из основных принципов Болонского процесса.

---

### **Причины проблем присоединения Украины к Болонскому процессу**

---

По нашему мнению, можно указать такие основные факторы, влияющие как на форму, так и на содержание обучения в условиях современной Украины. Первый из них – *государство*. В тоталитарных обществах этот фактор является определяющим. При проведении демократических реформ все большее значение приобретают еще два – *работодатель* со своими требованиями к необходимому ему содержанию образования и *студент (абитуриент)* со своими представлениями о перспективности избранной им специальности и о возможности того или иного университета удовлетворить его потребности. Эти факторы в современной Украине мало влияют на содержание образования в государственных университетах, но становятся немного заметными в частных учебных заведениях.

Есть еще один фактор, не связанный ни с тоталитарностью, ни с демократичностью общества – *корпоративные интересы преподавательской среды*.

Политика государства последовательностью не отличалась. В качестве подтверждения можно указать на весьма длительное сосуществование программ подготовки как магистров и бакалавров, так и специалистов. Напрашивающиеся же изменения в трудовом законодательстве, которые должны защитить права бакалавров, так и не осуществлены. Вторым примером может служить массовое представление лицензий частным учебным учреждениям и отделенным структурным подразделениям государственных университетов. Прошло совсем немного времени, и оказалось, что слишком большая их часть не в состоянии осуществлять образовательную деятельность. Третий пример – подписание документов о присоединении к Болонскому процессу. Взяв на себя тем самым серьезные обязательства, реальные шаги по приданию университетам автономии, а особенно по обеспечению академической мобильности студентов, отсутствуют. Игнорируется то, что для достижения упомянутой мобильности необходимо решить вопросы обновления материально-технической базы университетов, ремонта и увеличения площадей студенческих общежитий, модернизации инфраструктур студенческих городков, совершенствования визовой политики, рекламы нашего образования и, конечно, серьезного улучшения качества подготовки. Четвертый пример – принятие на уровне высших руководящих органов решений о содержании учебных планов (самым ярким является включение основ безопасности жизнедеятельности в учебные планы всех направлений).

В связи с этим, было проведено сопоставление наших учебных планов с планами 20 ведущих университетов Австрии, Великобритании, Франции, Германии, Италии, Греции, Польши, Болгарии, Сербии, Чехии, Литвы, Беларуси и Российской Федерации. Анализ показал, что учебные планы западноевропейских университетов содержат минимум (или совсем не содержат) непрофильных для конкретных специальностей дисциплин. Так, в университете Франца-Иосифа на специальности «Математика» непрофильных дисциплин нет совсем, а на специальности «История» как выборочные фигурируют «История экономики», «Историческая отраслевая информатика и документация» и «Организация производства и маркетинг для историков». В университетах Великобритании непрофильных дисциплин на большинстве специальностей нет. В Римском университете на химическом факультете как обязательная дисциплина фигурирует «Иностранный язык», на математическом - профильных дисциплин нет, на социологическом - как обязательные выступают «Статистика», «Современная история», «Английский язык», «Политическая экономия» и «Публичное право», по выбору идет «Право Европейского Союза». Дисциплины «Защита населения и хозяйственных объектов в чрезвычайных ситуациях», «Охрана труда» и «Основы Энергосохранения» замечены только в Беларуси как дополнительные виды обучения. Ближайшими по содержанию к нашим являются учебные планы Беларуси и Российской Федерации. Но их анализ показывает, что они находятся гораздо ближе к планам стран Европейского Союза, чем наши. Так, учебный план Российской Федерации по направлению «Геология» содержит семь названий в базовой части гуманитарных, социальных и экономических дисциплин и две в вариативной части. Следует подчеркнуть, что среди этих девяти названий тремя являются «Экономическая геология», «Правовые основы пользования недр» и «Менеджмент в сфере использования недр». Для сравнения укажу, что у нас таких дисциплин семнадцать.

Напрашивается вывод, что основным фактором, влияющим на формирование наших учебных планов, является четвертый из указанных ранее - корпоративные интересы преподавательской среды.

---

### **Исследовательские университеты – шанс для Украины**

---

Определенные надежды сейчас связываются с концепцией создания исследовательских университетов. Следует отметить, что еще совсем недавно даже политические лидеры Западной Европы недооценивали роль науки и образования. Как пишет академик М.З.Згуровский [6], наиболее образно эти взгляды высказал президент Франции в начале 70-х годов прошлого столетия Жорж Помпиду: «Самый простой способ потерять деньги – вложить их в сундук, женщин или в научные исследования». К чести Европы, она в начале нового тысячелетия приняла новую стратегию развития, суть которой сводится к повышению конкурентноспособности европейского рынка товаров и услуг, существенного увеличения количества

интеллектуально-наполненных рабочих мест за счет гармонизированного развития трех вершин так называемого «треугольника знаний», а именно – *образования, исследований и инноваций*. Ведущая роль в преодолении научно-технологического отставания Европы от других регионов мира возлагается на исследовательские университеты. Примеру Европы последовала и Российская Федерация. При разработке концепции исследовательских университетов в России исходили из того, что Национальный исследовательский университет (НИУ) – реальное воплощение нового подхода к качественной модернизации сектора науки и образования и новой институциональной формы организации научной и образовательной деятельности, призванной взять на себя основную нагрузку в кадровом и научном обеспечении запросов высокотехнологического сектора российской экономики. Исследовательский университет – высшее учебное заведение, одинаково эффективно осуществляющее образовательную и научную деятельность на основе принципов интеграции науки и образования. Важнейшими отличительными признаками НИУ являются способность как генерировать знания, так и обеспечивать эффективный трансфер технологий в экономику; проведение широкого спектра фундаментальных и прикладных исследований; наличие высокоэффективной системы подготовки магистров и кадров высшей квалификации, развитой системы программ переподготовки и повышения квалификации. Практически НИУ должен являться интегрированным научно-образовательным центром или включать ряд таких центров в виде совокупности структурных подразделений, осуществляющих проведение исследований по общему научному направлению и подготовку кадров для определенных высокотехнологичных секторов экономики. Стратегической миссией НИУ является содействие динамичному развитию научно-технологического комплекса страны и обеспечение его необходимыми человеческими ресурсами, сбалансированными по численности, направлениям подготовки, по квалификационной и возрастной структуре с учетом необходимых темпов их обновления и прогнозируемых структурных преобразований в науке и экономике. Основной задачей государственной поддержки института НИУ является вывод на мировой уровень образовательных организаций, способных взять на себя ответственность за сохранение и развитие кадрового потенциала науки, высоких технологий и профессионального образования, развитие и коммерциализацию в Российской Федерации высоких технологий.

В Украине совсем недавно были предприняты шаги по созданию исследовательских университетов. Правда, все создание превратилось в издание указов по присвоению соответствующего статуса университетам, часть из которых совершенно не отвечают западным представлениям об исследовательском университете. В уже упомянутой статье академик М.З. Згуровский отмечает, что многочисленные попытки реформирования высшей школы оставались и будут оставаться «косметическими», поскольку они никогда не имели целостного межотраслевого характера. А это задание выходит за рамки собственно образования. Украина нуждается в политической воле руководства страны, которое руководствовалось бы не сиюминутными выгодами для своей политической силы, а было бы способно планомерно проводить четкую долгосрочную политику в интересах всей страны.

---

## **Заключение**

В настоящее время «образовательное пространство» все в большей степени приобретает черты «сферы оказания образовательных услуг», что является следствием обретения образованием признаков производства с огромными финансовыми ресурсами. В Европе одним из основных инструментов создания единого образовательного пространства выступает «Болонский процесс», основанный на интеграционных тенденциях европейских стран и имеющий социальные, экономические и академические обстоятельства. Украина находится в начале «Болонского процесса» и нуждается, в первую очередь, в политической воле руководства страны. Одним из шансов интеграции Украины в общеевропейское образовательное пространство является создание исследовательских университетов, как единства образования, научных исследований и инноваций на основе еще сохранившегося в Украине высокого потенциала науки и образования.

---

## Благодарности

---

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

---

## Литература

---

- [1] [http://ru.wikipedia.org/wiki/Болонский\\_процесс](http://ru.wikipedia.org/wiki/Болонский_процесс)
  - [2] [http://ru.wikipedia.org/wiki/Студенческое\\_движение](http://ru.wikipedia.org/wiki/Студенческое_движение)
  - [3] [http://www.e-joe.ru/sod/98/4\\_98/st119.html](http://www.e-joe.ru/sod/98/4_98/st119.html)
  - [4] <http://southkoreanews.ru/story/921>
  - [5] <http://profedu.ru/news/element.php?ID=705>
  - [6] <http://www.zn.ua/3000/3300/54763/>
- 

## Информация об авторе

---



**Закусило Олег Каленикович** – первый проректор Киевского национального университета им. Т. Шевченко, член-корреспондент Национальной Академии педагогических наук Украины, доктор физ.- мат. наук, профессор, заведующий кафедрой исследования операций факультета кибернетики ;Украина, 01017, Киев, ул. Владимирская, 64; e-mail: [do@unicyb.kiev.ua](mailto:do@unicyb.kiev.ua),; [zakusylo@univ.net.ua](mailto:zakusylo@univ.net.ua)

**Сфера научных интересов:** исследование операций, теория вероятностей, математическая статистика, е-образование.

## ЗНАНИЕ-ОРИЕНТИРОВАННЫЕ СИСТЕМЫ РАЗРАБОТКИ ЭЛЕКТРОННЫХ КУРСОВ ОБУЧЕНИЯ

**Александр Палагин, Юрий Тихонов, Николай Петренко, Виталий Величко**

**Аннотация:** В работе рассматривается онтологический подход к проектированию и разработке электронных курсов. Приведен пример проектирования онтологии предметной дисциплины базы данных. Применение предложенного подхода позволит улучшить качество разрабатываемых электронных курсов, а использование средств автоматизированного лингво-семантического анализа текстовой информации на естественном языке даст возможность снизить трудозатраты на составление электронного курса.

**Ключевые слова:** онтологический подход, автоматизированная разработка электронного курса.

**ACM Classification Keywords** I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods, K.3 COMPUTERS AND EDUCATION - K.3.1 Computer Uses in Education

---

### Введение

Одним из перспективных направлений дальнейшего совершенствования электронных систем обучения является разработка методологических, онтологических и логических основ конструирования электронных курсов (ЭК). К онтологическим аспектам относится круг вопросов, начиная от областей применения и до формального описания компонент компьютерных онтологий предметных дисциплин (ПдД). Главный вектор исследований направлен на формализацию этапов построения, структурирования и представления знаний ПдД при обучении и интегрированного с ПдД проблемного пространства (эффективное сочетание лекционного и лабораторного материалов). В свою очередь, эффективная реализация указанных этапов и получение конечного результата (в виде, библиотеки онтологических баз знаний предметных дисциплин) невозможно без проведения системно-онтологического анализа заданной совокупности предметных дисциплин [1].

В данной работе будет рассмотрен онтологический аспект проектирования электронных курсов (ЭК).

---

### Постановка задачи

Онтологический поход (ОнП) к проектированию и разработке ЭК является одним из важных практических приложений направления онтологического инжиниринга, с помощью которых могут быть решены (и качественно улучшены решения) следующие актуальные задачи в области электронного (в том числе и дистанционного) обучения:

- автоматизированная разработка ЭК на основе лингво-семантического анализа больших объемов текстовой информации с помощью оригинальных инструментальных средств. При этом текстовая информация формируется из апробированных в ведущих ВУЗах страны оцифрованных учебников (ОУ) на украинском и русском языках по данной ПдД;
- структурирование и обработка терминов и понятий, содержащихся в ОУ в соответствии с утвержденной методологией по заданной ПдД;
- существенное уменьшение трудоемкости составления ЭК, в т.ч. новых ЭК.

---

### Онтологический подход

В онтолого-классификационной схеме средств и методов искусственного интеллекта ОнП [2] трактуется как разновидность системного подхода, основанного на знаниях. ОнП обеспечивает эффективное проектирование компонент любой знание-ориентированной информационной системы (ЗОИС).



Полагаем что, знание-ориентированная система разработки ЭК (ЗОЭК), как ЗОИС с обработкой знаний, содержащихся в оцифрованных учебниках, разработана с учетом *онтологического подхода*, если она обладает следующими характерными чертами:

1. Компьютерные онтологии (КО) ПдД обеспечивают эффективную машинную обработку общезыковых и предметных знаний ПдД.
2. В отличие от обычного, субъективного подхода при разработке ЭК, системно-онтологический подход [1] предполагает строгую (насколько это возможно на данном этапе развития науки) структуризацию терминов и понятий ПдД. Категориальный уровень представляется *онтологией верхнего уровня* домена предметных дисциплин ВУЗа (ОВУ). Проектирование ОВУ должно быть включено в общий алгоритм разработки ЭК. Например, на кафедре информационных технологий и систем Луганского национального университета ПдД для различных специальностей связаны через ОВУ.
3. Архитектурно-технологические особенности:
  - модельно-управляемая архитектура характеризуется высоким уровнем формальности представления онтологии ПдД и механизмами онтолого-управляемости;
  - высокая степень интеграции предметных знаний по совокупности дисциплин кафедры.
4. Использование оригинальных инструментальных средств поддержки автоматизированного построения онтологий ПдД (методика, технология и программная реализация).

Общая архитектурно-структурная организация онтолого-управляемой информационной системы описана в [2]. На рис. 1 представлена схема фрагмента архитектуры, отображающая “онтологическую” компоненту ЗОЭК, а также связь с процедурами информационной модели обработки знаний, содержащихся в оцифрованных учебниках.

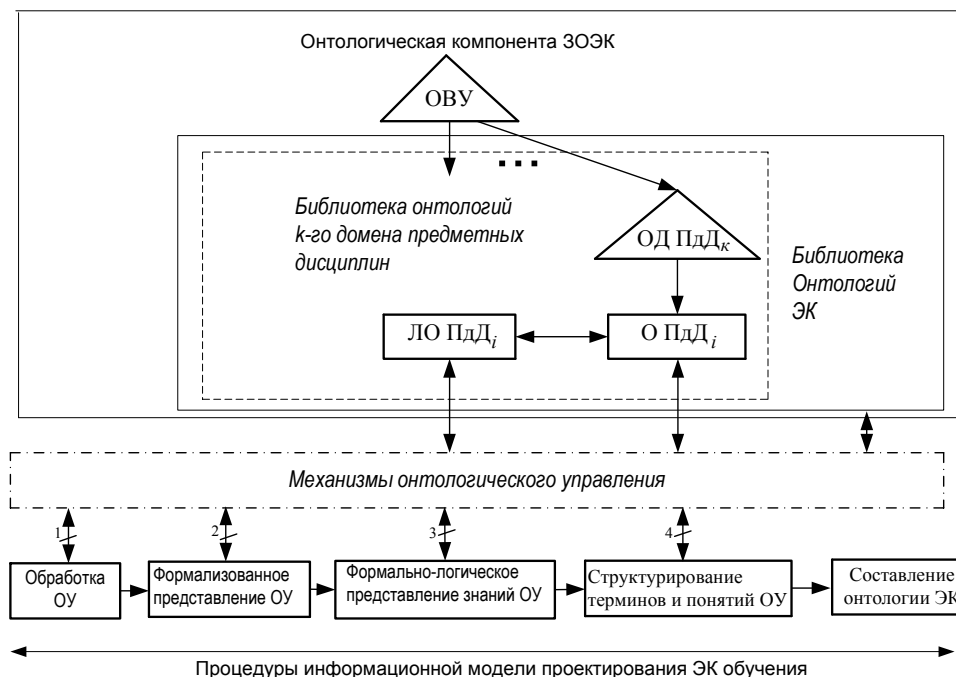


Рис. 1. Фрагмент архитектуры “онтологической” части ЗОЭК

На рис. 1 приняты следующие обозначения: ОЭК – онтологизированные электронные курсы; ОУ – оцифрованный учебник; ОД ПдД<sub>к</sub> – онтология *k*-го домена предметных дисциплин факультета, где  $k = \overline{1, K}$ ,  $K = \text{Card } D$  – мощность множества онтологий доменов *D*, интегрированных в библиотеку; ЛО ПдД<sub>і</sub>

– лингвистическая онтология  $i$ -ой предметной дисциплины, где  $i = \overline{1, l}$ ,  $l = \text{Card ПдД}$  – мощность множества онтологий предметных областей, интегрированных в библиотеку;  $O_{\text{ПдД}_i}$  – онтология  $i$ -ой предметной дисциплины,  $LO_{\text{ПдД}_i}$  – лингвистическая онтология  $i$ -ой предметной дисциплины, где  $i = \overline{1, l}$ ,  $l = \text{Card ПдД}$ ; 1 – обработка омонимичных форм; 2 – построение поверхностных семантических структур; 3 – обработка семантических неоднозначностей; 4 – обработка концептуальных структур.

---

### Компьютерные онтологии

---

Компьютерные онтологии – это интенсивно развивающийся раздел информатики, как теоретической, так и практической [1, 3, 4]. Актуальность данного направления в настоящее время очевидна.

КО некоторой ПдД можно рассматривать как общезначимую, открытую базу знаний ПдД, представленную на общепринятом (формальном) языке спецификации знаний. Это позволяет выделить из множества разнообразных практических приложений КО [5, 6] приложение с особым статусом – обучение. При этом создание “онтологизированных” электронных курсов (в том числе и дистанционных) представляется важной самостоятельной научно-технической проблемой.

Только информационная система с онтолого-управляемой архитектурой, имеющая в своём составе КО некоторой ПдД, позволяет реализовать все преимущества электронного образования, в том числе и дистанционного обучения. Наиболее полно указанные свойства проявляются при использовании и взаимодействии онтологий двух уровней – онтология домена предметных дисциплин (множество дисциплин, читаемых на факультете) и онтологий самих предметных дисциплин. При этом естественным образом находят своё решение такие проблемы как: открытости и закрытости систем обучения; единообразия представления понятийных структур обучения; автоматизированного построения новых курсов обучения с учётом накопленных преподавателями кафедры знаний и др.[7].

Добротный ЭК должен обладать всеми структурными элементами определяющими аксиоматизацию научных знаний: аксиомы или схемы аксиом; исходные (первоначальные) понятия; правила вывода и построения системы знаний; определения; теоремы, следствия, положения, выводимые из теории, леммы, логико-методологические принципы аксиоматизирующей теории, *интерпретация* как логический прием по установлению значения терминологических выражений теорий, знаковой (формальной) системы [8].

Произвольная онтология (простая, смешанная или формальная) эффективна, прежде всего, в обучении. Действительно, гораздо эффективнее показать студенту онтограф ПдД, чем другими средствами описывать концепты ПдД и их связи.

Онтология обладает когнитивностью, структурирует знания некоторой предметной области в соответствии с онтологическими принципами. Преподавателю онтология (в особенности компьютерная) облегчит разработку учебного курса, студенту улучшит восприятие материала и взаимосвязи предметных знаний. В силу этого можно ожидать значительного эффекта от применения онтологий в образовании в виде улучшения качества восприятия знаний в ПдД.

Ожидается, что онтологии будут влиять на следующее поколение систем e-learning.

Как уже говорилось, наиболее полно преимущества онтолого-управляемой информационной системы в ВУЗе проявляются при использовании и взаимодействии онтологий двух уровней – онтология домена предметных дисциплин (множество дисциплин, читаемых на кафедре) и онтологий самих предметных дисциплин.

Онтология в реальных ЗОЭК, обеспечивающих работу кафедры ВУЗа, должна включать две иерархически связанных компоненты: метаонтологию (онтологию, концептами которой служат онтологии последующих категориальных уровней), предметную онтологию (состоящую из онтологии объектов и онтологии процессов и задач).

## Модельно-управляемая архитектура

Использование онтологии в качестве компонента интерфейса пользователя особенно важно при автоматизированном создании ЭК, так как онтология включает семантическую информацию относительно ограничений, наложенных на понятия и их отношения. Онтологии, например, могут быть применены для генерации интерфейсов для ЭК, которые проверяют нарушение ограничений. Преподаватель и студент могут ознакомиться с онтологией, чтобы лучше понять терминологический словарь, используемый в ПдД.

Реализация указанных технологий предполагает учет различных формально-методологических требований, критериев и оценок. Приведем основные из них.

1. Сущность методов онтологического исследования как раздела системного анализа предполагает:

- построение информационной и функциональной моделей ЗОЭК;
- исследование структур ПдД и взаимосвязи компонент знаний для ПдД, читаемых на кафедре.

2. Необходимость структурирования терминов и понятий.

3. Правила формирования достоверных утверждений и выводов, описывающих термины и понятия ПдД.

4. Поддержание онтологии домена ПдД, читаемых на кафедре.

## Средства инструментальной поддержки автоматизированного построения О ПдД

Одним из главных требований к онтологизированным (знание – ориентированным) системам разработки ЭК (с формированием терминов и понятий из ЕЯО ОУ) является наличие формализованной методики автоматизированного построения онтологии произвольной ПдД, действенной информационной и функциональной модели и соответствующих инструментальных средств программной реализации. При этом предполагается, что в качестве источника онтологических знаний выступает лингвистический корпус текстов множества ОУ (на русском и украинском языке) из заданной ПдД.

Для обработки текстов на украинском языке известны только лингвистические программные процессоры, использующиеся в основном для построения лексикографических систем. Кроме того, отдельной проблемой для всех систем-прототипов представляется лингвистическая обработка больших объемов текстовой информации во времени, приближенном к реальному.

Для модельно-управляемой ЗОЭК с обработкой знаний, содержащихся в ОУ (рис.1.), схема функциональной модели процесса проектирования ЭК представлена на рис. 2.

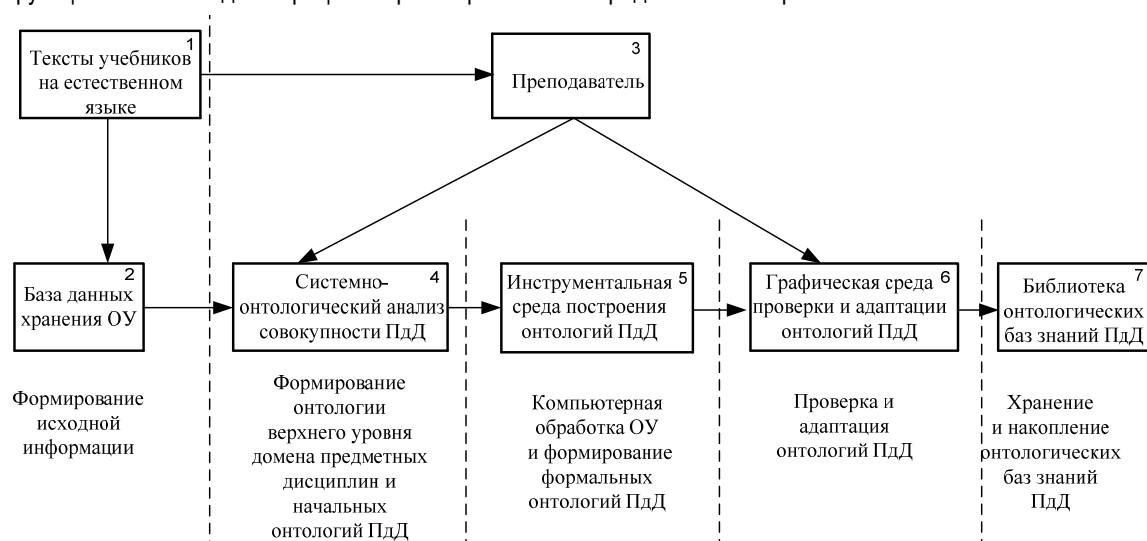


Рис.2. Функциональная модель процесса проектирования ЭК обучения.

Процесс проектирования ЭК ориентирован на эффективную программную поддержку реализации наиболее трудоёмких по времени процедур, что должно решить проблему обработки больших объемов текстовой информации (см. блоки 1, 2).

В блоке 4 выполняется системно-онтологический анализ совокупности ПдД и формирование ОВУ.

В блоке 5 выполняется компьютерная обработка большого объема текстовой информации ОУ и формируется онтология ПдД. Построение множества понятий ПдД является первым шагом в проектировании ЭК. Известны многочисленные публикации [4-6, 9-11 и др.], описывающие ручные и “автоматизированные” методики проектирования онтологии ПдО. Их общим недостатком является ручное формирование множества понятий. При объеме понятий для ЭК из нескольких ОУ в несколько тысяч такое проектирование становится крайне неэффективным. ЗОЭК поможет устранить данный недостаток.

В блоке 6 осуществляется проверка и адаптация онтологии ПдД в интерактивном режиме, включающем работу преподавателя в автоматизированной графической среде.

В блоке 7 осуществляется накопление онтологических баз знаний ПдД, использование которых позволит обеспечить единообразие представления понятийных структур обучения дисциплин читаемых на кафедре. Электронный курс, построенный на основе онтологического подхода автоматизированным способом по данной ПдД, интегрирует общие концепты и их связи и в этом смысле является общезначимым, общепринятым, построенным с существенным уменьшением субъективного фактора.

Построенная в соответствии с функциональной моделью программная модель позволит существенно уменьшить трудозатраты на составление ЭК (в т.ч. новых ЭК) за счет большой степени автоматизации работ, выполняемых в блоке 5 (рис.2.).

Интегрированные инструментальные средства, реализующие ЗОЭК, находятся в стадии проектирования и разработки.

ЭК, разработанные с использованием онтологического подхода, позволят реализовать большинство преимуществ электронного образования. Знания выпускника при использовании онтологизированных ЭК будут представляться с помощью единой системы однозначно трактуемых понятий, при этом систему понятий легче сконструировать непротиворечивой, полной, структурированной, адекватно отражающей действительность.

---

### **Пример проектирования онтологии ПдД “Базы данных”**

---

Принципы и механизмы онтологического подхода применяются при проектировании архитектурной и информационной компонент КС. Далее рассмотрим пример проектирования информационной компоненты ЗОЭК с обработкой ОУ – онтологии ПдД “Базы данных” (БД). При этом, методика проектирования (и соответствующий алгоритм) описана в [1], а базовой компонентой инструментальных средств обработки множества ОУ является языково-онтологическая информационная система, которая состоит из лексикографической базы знаний, языково-онтологической картины мира, семантического анализатора и семантико-информационного интерпретатора [12].

В качестве множества ЕЯО использовались ОУ по БД [13-19], На рис. 3 представлен онтограф начальной онтологии заданной ПдД, который в соответствии с методикой спроектирован вручную преподавателем кафедры.

На рис. 4,а–4,б представлен онтограф онтологии объектов заданной ПдО, спроектированный с помощью оригинальных инструментальных средств.

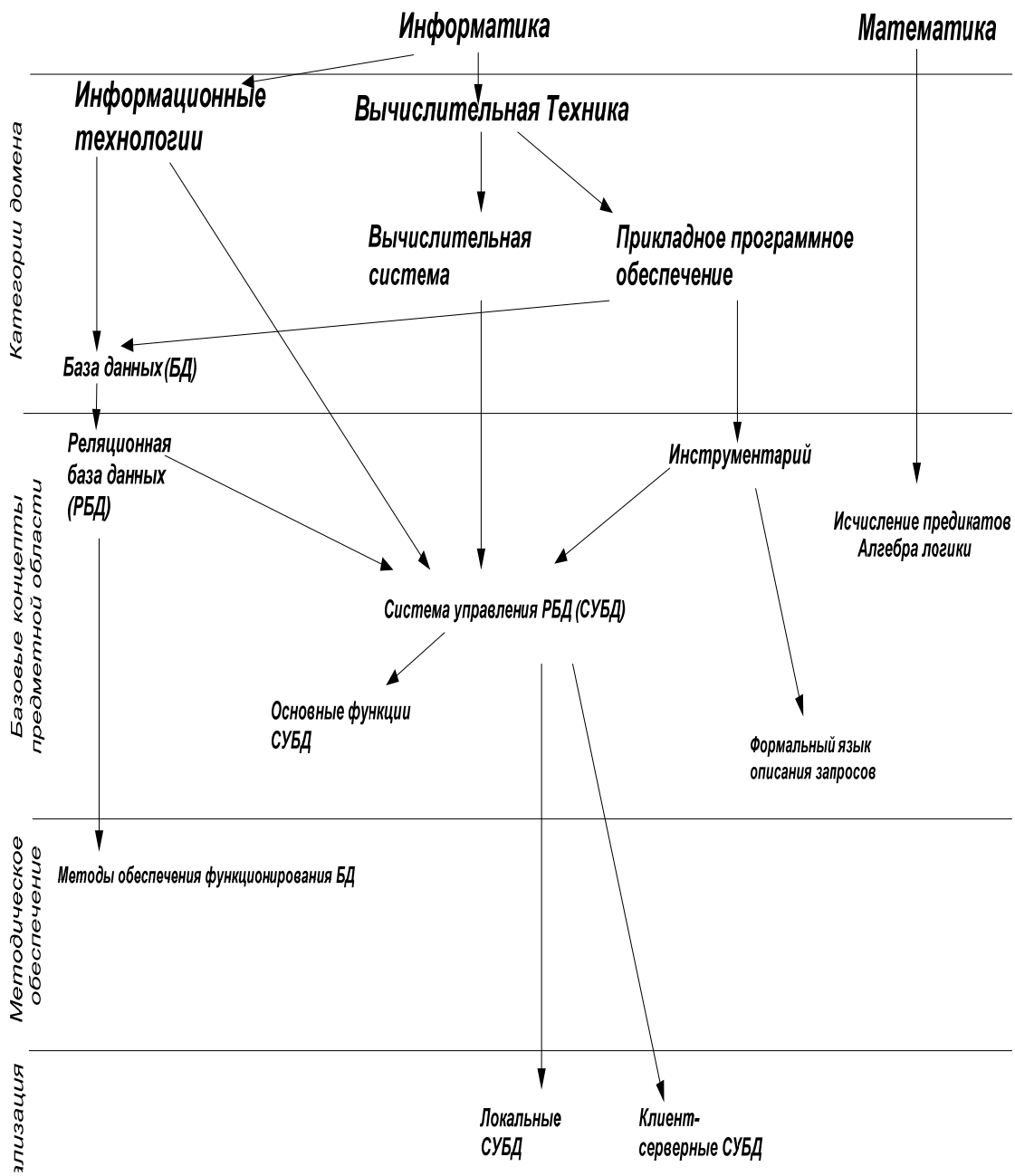


Рис. 3. Онтограф начальной онтологии ПдД "БД "

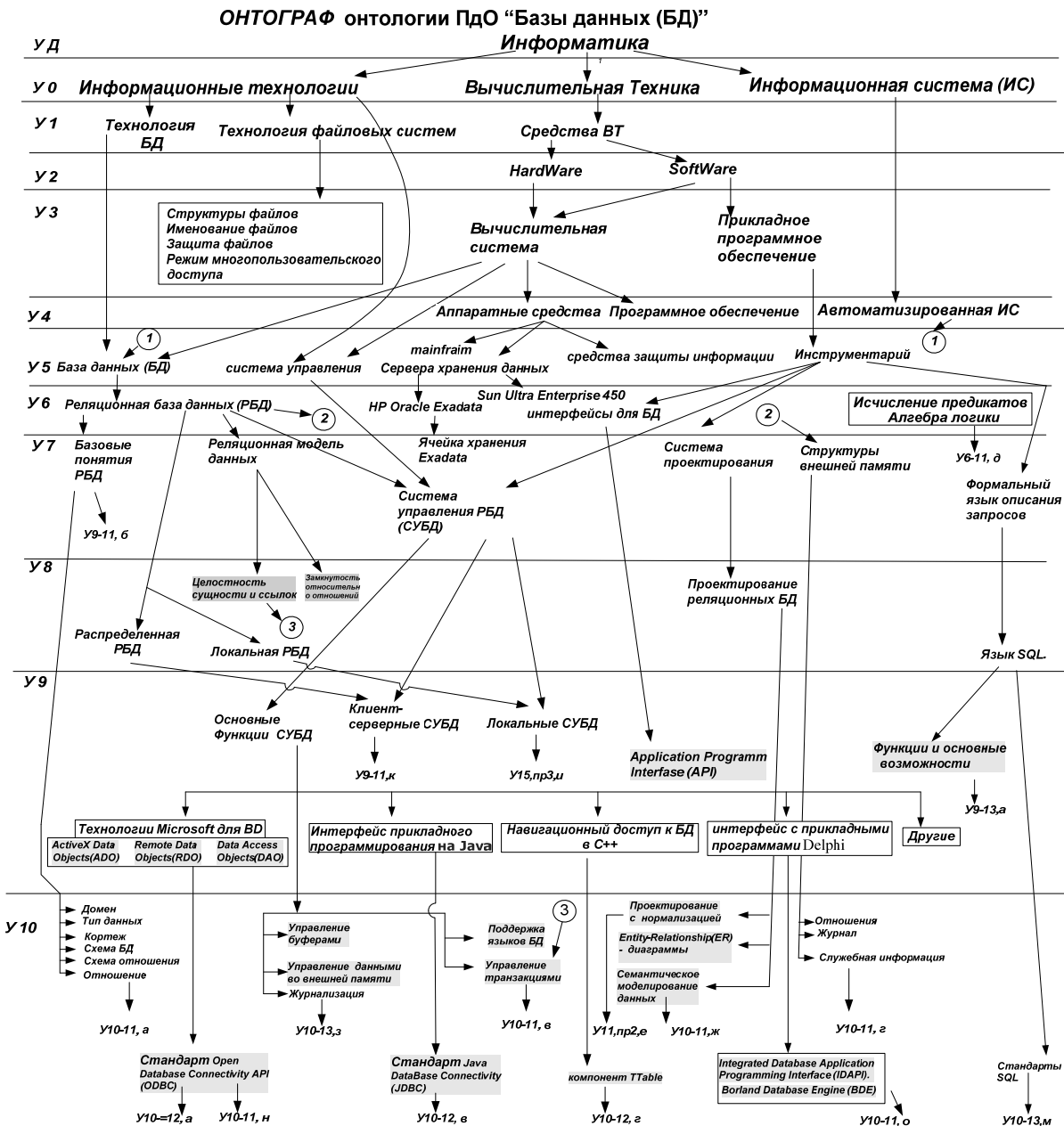
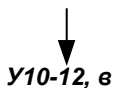


Рис. 4, а. Онтограф онтологии объектов ПдД "БД"

На рис. 4,а-4,б введены обозначения:



– ссылка на объект в этом рисунке;



– ссылка на объект в другом уровне на другом рисунке, **у10-12** указывает с какого уровня на какой указывает ссылка, **в-** индекс ссылки.

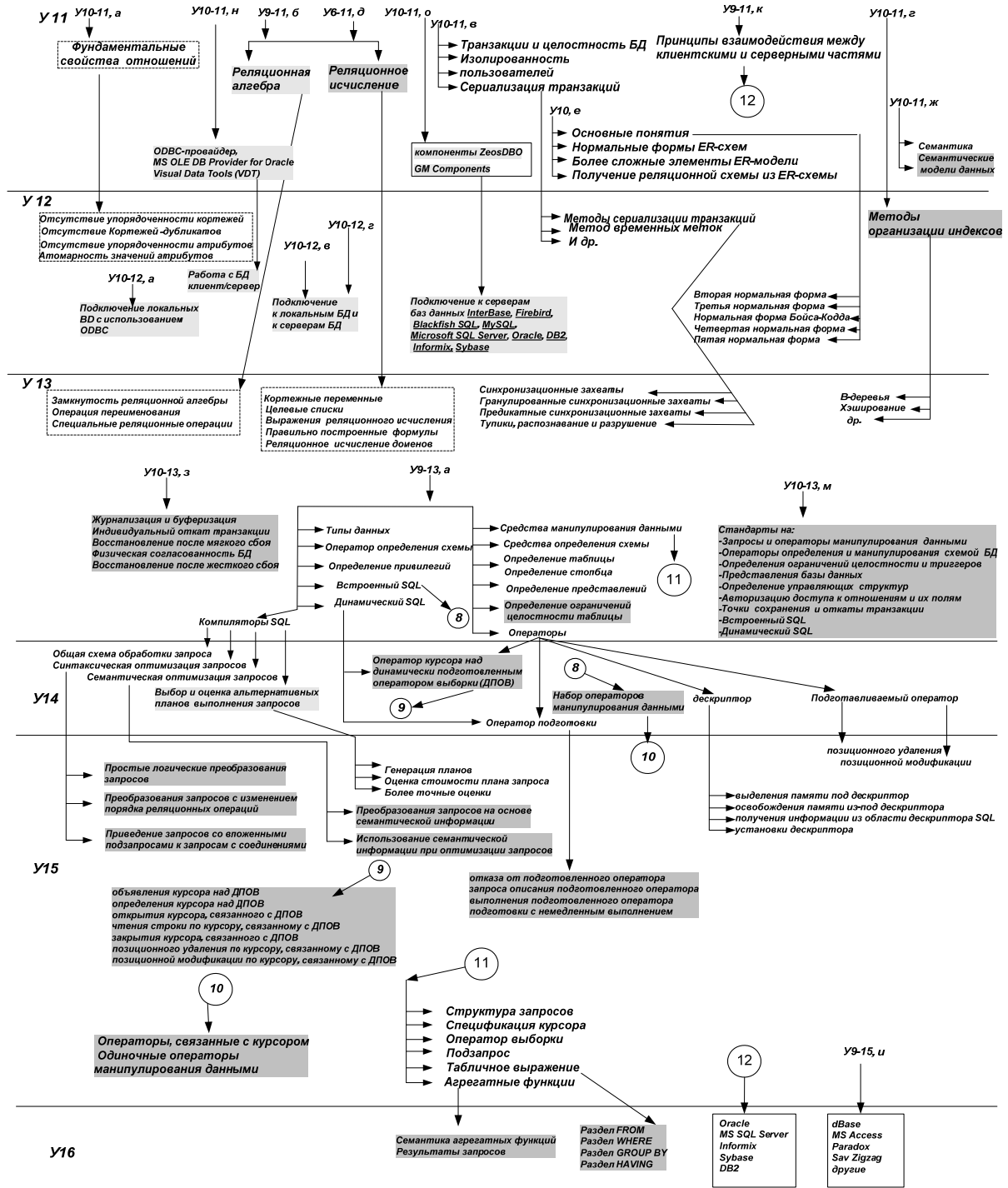


Рис. 4, б. Онтограф онтологии объектов ПдД "БД "

---

## Выводы

---

В работе предложен подход к решению актуальных задач в области электронного образования, в частности автоматизированной разработки ЭК на основе онтологического подхода. Использование в процессе проектирования ЭК автоматизированного лингво-семантического анализа больших объемов текстовой информации на естественном языке, полученной из ОУ на русском и/или украинском языке, а также структурирование и обработка терминов (понятий) из заданной ПдД, позволяет существенно улучшить качество и снизить трудозатраты на составление ЭК. Совершенствование предлагаемого подхода предполагается на основе опытной эксплуатации системы поддержки автоматизированного построения О ПдД.

---

## Литература

---

1. Палагин А.В., Петренко Н.Г. Системно-онтологический анализ предметной области // УСИМ. – 2009. – № 4. – С.3–14.
2. Палагин А.В., Крывый С.Л., Петренко Н.Г. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация // УСИМ. – 2009. – № 3. – С.42–55.
3. Gruber, T.R. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies*, 43(5/6): 907–928.
4. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
5. Палагин А.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: «УНІВЕРСУМ-Вінниця», 2005. – 680 с.
6. Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии: Учеб. пособие. – М.: Изд. – во МГТУ им. Н.Э. Баумана, 2005. – 304 с.
7. Палагин А.В., Петренко Н.Г., Тихонов Ю.Л., Величко В.Ю. К вопросу автоматизированного построения онтологии предметной дисциплины для электронных курсов обучения. *Вісник східноукраїнського національного університету імені Володимира Даля* – 2010. – № 4(150).
8. Добров Б.В., Лукашевич Н.В., Невзорова О.А., Федунев Б.Е. Методы и средства автоматизированного проектирования прикладной онтологии // *Изв. РАН. Теория и системы управления*. – М.: 2004. – № 2. – С.58–68.
9. Gruber, T. R. 1993. A translation approach to portable ontology specifications // *Know-ledge Acquisition*, 5 (2): pp. 199–220.
10. Гаврилова Т.А. – Онтологический инжиниринг / [http://www.kmtec.ru/publications/library/authors/ontolog\\_engeneering.shtml](http://www.kmtec.ru/publications/library/authors/ontolog_engeneering.shtml) .
11. Natalya F. Noy and Deborah L. McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. – available at [http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html).
12. Палагин А.В., Петренко Н.Г. К проектированию онтологууправляемой информационной системы с обработкой естественно-языковых объектов // *Математические машины и системы*. – 2008. – №2. – с.14–23.
13. Д.Дейт. Введение в систему баз данных . - М., СПб.: BHV - Санкт - Петербург 1977.- 312с
14. Скотт Баркер. Использование Microsoft Access 97.-Киев-Москва: Диалектика, 1997г.- 257 с.
15. Мари Свонсон. Microsoft Access 97: наглядно и конкретно.-Москва: Microsoft Press, Русская редакция, 1997г.- С10-57.
16. Алексей Гончаров. Access 97 в примерах.-С.-Петербург: Питер, 1998г.-С.12-60.
17. В. Пасько. Access97 для пользователя.-Киев: BHV, 1997 г.-С.5-28, 99-121
18. Фаронов В. Программирование баз данных в Delphi 6. Учебный курс. – СПб: Питер, 2002. – 352 с.
19. Шумаков П.В. Delphi 3 и разработка приложений баз данных. – М.: «Нолидж», 1998. – 704 с.



---

**Информация об авторах**

---



**Палагин Александр Васильевич** – академик НАН Украины; заместитель директора Института кибернетики имени В.М. Глушкова НАН Украины; Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: palagin\_a@ukr.net

Основные области научных исследований: системная интеграция трансдисциплинарных научных знаний, онтологический инжиниринг



**Тихонов Юрий Леонтьевич** – к.т.н., доц., доцент кафедры ИТС, Луганского национального университета им. Тараса Шевченко; e-mail: t2003i@mail.ru

Основные области научных исследований: разработка онтологизированных электронных курсов



**Петренко Николай Григорьевич** – к.т.н., старший научный сотрудник, Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: petrng@ukr.net

Основные области научных исследований: методология и инструментальные средства автоматизированного проектирования онтологий предметных областей, системная интеграция трансдисциплинарных научных знаний



**Величко Виталий Юрьевич** – к.т.н., доц., Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: velychko@aduis.com.ua

Основные области научных исследований: индуктивный логический вывод, обработка естественно-языковых текстов.

## ДИДАКТИЧЕСКИЙ ДИЗАЙН СИСТЕМЫ ОБУЧЕНИЯ НОВОГО ПОКОЛЕНИЯ

Елена Вахтина, Александр Вострухин

**Аннотация:** В статье рассматривается дидактический дизайн как проектная научно-методическая деятельность педагога по разработке системы обучения нового поколения, которая ориентирована на формирование среды опережающей подготовки инженеров. Дидактическая система (D-система) определяется ведущей в развитии образовательной среды университета, на основании того, что является непосредственным исполнителем, осуществляющим профессиональную подготовку инженеров. Приводятся требования к D-системе нового поколения. Показана модель развития D-системы, в которой выделены фазы: моделирования, проектирования, конструирования и эксплуатации. На примере междисциплинарного учебного модуля «Микропроцессорная техника» представлен опыт разработки и внедрения в учебную практику образовательного ресурса, удовлетворяющего требованию опережающей подготовки инженера. Образовательный ресурс состоит из контента, изложенного в учебном пособии, и инструментария – стенда микроконтроллерного. Контент представлен в форме инновационной практико-ориентированной задачи с квалифицированным дидактическим сопровождением этапов ее решения. Стенд микроконтроллерный – в форме промышленного образца лабораторного стенда.

**Ключевые слова:** дидактический дизайн, дидактическая система, опережающая подготовка, междисциплинарный учебный модуль, микроконтроллер.

**ACM Classification Keywords:** J.2 Physical Sciences and Engineering – Electronics, Engineering, 1.2.6 Learning – Concept learning.

---

### Введение

В настоящее время инженерное образование представляет собой самую масштабную подсистему высшего профессионального образования в России и за рубежом. Наблюдается тенденция повышения спроса у молодежи на инженерные специальности. Однако, несмотря на масштабность подготовки инженеров, остается дефицит специалистов в области наукоемких и интеллектуальноемких технологий. По прогнозам экспертов США и Японии этот дефицит будет расти. Ведущие страны мира бьют тревогу по поводу того, что национальные системы профессионального образования не удовлетворяют в полной мере спрос на инженерные кадры на рынке интеллектуального труда как в количественном, так и, в особенности, качественном отношении [Похолков, 2000].

Проблемой исследования мы определили выявление педагогических условий формирования готовности будущих инженеров к продуктивной и инновационной деятельности в образовательном процессе вуза.

Для ее решения мы обратились к новому феномену культуры – дизайну. «Дизайн, являясь органичной частью современной культуры, особенно рельефно подчеркивает ее проектность. Как правило, дизайнер выполняет несколько профессиональных ролей. Во-первых, он выступает как исследователь и тогда действует в соответствии с нормами научно-теоретической деятельности. Во-вторых, ему приходится выполнять функции инженера-проектировщика и методиста, рассматривая продукт своей деятельности как особого рода проект. В-третьих, он является художником, наследующим и эстетически преобразующим все достижения предшествующей художественной культуры в целях создания нового произведения искусства» [Горохов, 2007]. «Вот эта многоликость и единство профессиональных ролей выводят дизайн на ведущее место в выборе методологической платформы реализации компетентностной модели в инженерном образовании» [Соснин, 2009].

Сформировавшиеся направления дизайна: промышленный, архитектурный, ландшафтный, текстильный и др. дополняется в настоящее время новыми направлениями, одним из которых можно назвать дизайн дидактический (В.Э. Штейнберг, Н.В. Соснин). Объектом исследования этого научного направления -

является проектирование гармоничной образовательной среды в социальных системах различного масштаба. Предметом – научно обоснованные проекты дидактических систем, процессов и объектов.

Под дидактическим дизайном (D-дизайном) мы понимаем проектную научно-методическую деятельность педагога по разработке инновационных программ обучения и формированию гармоничной образовательной среды.

В педагогике сложился так называемый средовой подход, который можно рассматривать не только как концептуальное направление, но и как технологию опосредованного управления процессом формирования и развития личности (В.А. Караковский, Л.И. Новикова, Н.Л. Селиванова и др.).

---

### **D-дизайн системы обучения нового поколения**

---

Очевидно, что для подготовки способных к созданию инновационных продуктов инженеров, образовательная среда университета должна быть все время в развитии. В исследованиях Н.К. Нуриева, Л.Н. Журбенко, С.Д. Старыгина и др. показано, что рамки требований к выпускникам-инженерам, подготовленным в образовательной среде университета, находятся в следующих пределах: минимальный (репродуктивный) – выпускник должен удовлетворить потребности экономики социума на актуальный момент, а максимальный (продуктивный) – являться “творцом” развития общества в определенном направлении инженерии. Поэтому состояние образовательной среды университета во многом определяет развитие общества в целом. При этом дидактическая система (D-система) является “локомотивом” развития образовательной среды университета, т.к. непосредственно осуществляет подготовку инженеров необходимых обществу.

D-система нового поколения должна удовлетворять минимум четырем требованиям [Нуриев, 2009].

*Требование первое.* D-система должна быть открытой развивающейся системой, т.е. в находиться в проектировании перехода с достигнутого уровня (прототип-1) на новый перспективный (прототип-2), определяемый изменениями в требованиях к подготовке инженеров. Другими словами в проекте D-системы должна быть заложена программа развития самой системы.

Одним из предлагаемых «сверху» механизмов обновления D-систем является новое поколение образовательных стандартов ФГОС ВПО, вводимых в действие с 2010 г., которые имеют свои преимущества и недостатки.

Преимущества ФГОС ВПО:

- задает требования к качеству подготовки инженера, детерминирующие изменения D-системы и определяющие ее рамки (формат);
- направлен на переориентацию среды обучения с предметоцентрированной на студентоцентрированную (результатом образования должны стать не столько дисциплинарные знания, сколько способы мышления и деятельности);
- не содержит ограничений тормозящих развитие D-систем.

Недостатки:

- описательный характер требований ФГОС ВПО позволяет их трактовать по-разному, оставляя простор для субъективных оценок как исполняющих, так и контролирующих эти требования;
- темпы создания ФГОС ВПО отстают от темпов развития инженерии особенно в тех областях деятельности, которые развиваются бурными темпами, такими как, например, информационные технологии и электроника, нанотехнологии и т.п.;
- в стандартах не учтено то обстоятельство, что ориентация на компетентностную модель образования не только не исключает, но, наоборот, требует использования знаниевого подхода. Так, например, в основной образовательной программе новой редакции ФГОС ВПО для электротехнических направлений подготовки не нашлось места «Теоретическим основам электротехники» (ТОЭ). Эта дисциплина фундаментальна для всего электротехнического

образования. На ней базируется оборонная электротехническая тематика и электромеханика. Эта дисциплина не имеет зарубежных аналогов [Бутырин, 2009]. И, наверное, поэтому она была исключена из стандартов нового поколения, ориентированных на болонский процесс, порожденный интеграцией европейского образовательного пространства. Однако интеграция не исключает, а наоборот предполагает использование достижений национальных культур как фактора развития единого образовательного пространства.

- не учитывается отечественный передовой опыт подготовки инженеров (МФТИ, мехмат МГУ, МГТУ им. Н.Э. Баумана, ТПУ и др.), широко представленный в публикациях наиболее востребованных научно-педагогических журналов: «Образовательные технологии и общество», «Высшее образование в России», «Высшее образование сегодня» и др.

Эти недостатки придется восполнять «снизу» кафедрой при условии ее опережающего развития, которое обеспечивается обратной связью профессорско-преподавательского состава (ППС) кафедры с проблемной средой в профессиональной деятельности.

Развитие образовательной среды университета происходит за счет проектирования новых прототипов D-систем следующим образом. Подготовка инженера происходит путем обучения системе дисциплин, определенных стандартом. Организуется образовательная деятельность в рамках каждой из них. Таким образом, каждая учебная дисциплина является компонентом среды обучения и средством достижения цели-результата, выраженного через требуемый уровень подготовки, который периодически корректируется новым поколением стандартов. Следовательно, и сама учебная дисциплина должна быть все время в развитии, которое осуществляется через анализ результатов обучения, прогноз изменений в содержании и технологии обучения, проектировании нового прототипа D-системы. В этом процессе важную роль играют исследования и прикладные разработки ППС кафедры в проблемной среде профессиональной деятельности. Они актуализируют содержание преподаваемых дисциплин и мотивируют активность студентов.

Развитие учебной дисциплины в рамках направления подготовки происходит по спиральной схеме, в которой выделяются фазы моделирования, проектирования, конструирования и эксплуатации (рис. 1).



Рис. 1. Модель развития D-системы

В первой фазе (моделирование) – на основе изменений в условиях и факторах инженерной деятельности, и появлением новых требований к подготовке инженеров прогнозируются изменения компонентов действующего прототипа-1 D-системы, и создается модель нового прототипа-2. В фазе проектирования разрабатываются содержательный и процессуальный (технологический) блоки учебной дисциплины в рамках нового прототипа-2. Фаза конструирования предусматривает проверку (апробацию) прототипа-2 (дидактического проекта учебной дисциплины) и его корректировку на основе полученных экспериментальных данных. Фаза эксплуатации состоит из двух этапов: внедрения и работы прототипа-2 D-системы в образовательном процессе.

*Требование второе.* В современных условиях полипарадигмальности D-система должна базироваться на системе подходов: знаниевом, деятельностном и компетентностном. Эти подходы, развивающиеся в значительной степени независимо и имеющие разные целевые установки, в системном взаимодействии обеспечивают целостное развитие личности студента.

Как отмечают многие отечественные и зарубежные исследователи, создание студентоцентрированной среды обучения достигается через полновесную реализацию компетентностного подхода. Именно в рамках этого подхода смещаются акценты в оценке результатов обучения с входных факторов на развитие выходных характеристик, т.е. востребованных на рынке труда знаний, умений и компетенций. Это ведет к изменению роли преподавателя и в связи с этим к необходимости освоения им новых компетенций, иной организации обучения, которая начинает определяться запросами студентов и должна обеспечить большую их вовлеченность и самостоятельность.

Под компетентностью понимают «способность применять знания, умения и личностные качества в определенной области деятельности», но прежде чем применять, нужно сформировать эти знания и умения. Для этого и сегодня, и в будущем останутся востребованными освоенные образовательной практикой знаниевый и деятельностные подходы.

*Требование третье.* D-система нового поколения должна быть системой с виртуальной составляющей или реально-виртуальной дидактической системой (RVD-системой).

С появлением Интернет, Интранет, сотовой сети и широким распространением на все сферы жизнедеятельности информационно-коммуникационных технологий D-система, как компонент системы образования современного общества, не могла не использовать принципиально новые возможности организации учебного процесса в виртуальной среде. Наряду с традиционной (классической) организационной формой учебного процесса сегодня существует дистанционная форма с использованием сетевой системы управления обучением (Learning Management Systems – LMS), реализующей концепцию виртуальных учебных сред (Virtual Learning Environments), а также различные комбинации этих форм (blended learning – смешанное обучение). Последний вариант наиболее предпочтителен, так как сочетает достоинства обеих форм и компенсирует недостатки одной, за счет достоинств другой формы, тем самым, расширяя дидактические возможности D-системы.

Формирование виртуальной составляющей D-системы идет поэтапно. Первоначально имеющийся учебно-методический комплекс по дисциплине перерабатывается с бумажного в электронный вариант. При этом решаются вопросы эргономики: обрабатываются тексты с целью улучшения их восприятия, зонирование информации на экране, решаются вопросы психофизиологии цвета, проектируются статические и динамические иллюстрации и пр. Затем выбираются специальные инструментальные средства и оболочка для конструирования электронного образовательного ресурса (ЭОР). В практике инженерной подготовки сложился определенный набор элементов ЭОР: электронный учебник, виртуальный лабораторный практикум, задачи и упражнения, контрольные тесты, глоссарий, персоналии, хрестоматийные статьи и др. Следует отметить, что разработка содержательного наполнения образовательного ресурса очень трудоемкая и кропотливая работа, которая занимает у преподавателей много времени. Использование специальных инструментальных средств и оболочек экономит время разработки, обеспечивает современный уровень функциональных и коммуникационных возможностей и

пользовательского графического интерфейса, исключает многие ошибки начинающих разработчиков. Поэтому выбор конкретной системы программной оболочки очень важен.

В качестве основных параметров выбора могут быть предложены следующие показатели LMS:

1. Набор свойств и функций, удовлетворяющий потребности учебного учреждения в организации обучения.
2. Стоимость приобретения, использования или разработки системы, т.е. полная стоимость владения системой.
3. Возможность работы с национальным алфавитом.
4. Производительность, платформа, требования к аппаратному и программному обеспечению сервера.

Многие зарубежные системы управления обучением обладают великолепной функциональностью, производительностью, но имеют проблемы с отображением символов национального алфавита. К сожалению, несмотря на хорошие потребительские качества таких систем, эти проблемы не позволят использовать их в реальном учебном процессе.

5. Поддержка стандартов Интернет и дистанционного обучения.

Данный показатель важен для переносимости курсов, тестов, учебных планов и программ, разработанных в разных системах управления учебным процессом. Совместимость с образовательными стандартами и спецификациями позволит осуществлять композицию-декомпозицию образовательных ресурсов, поиск, обмен пользователями, кредитами, системами доставки и поддержки между разными образовательными организациями.

Хотелось бы особо остановиться на вызывающей уважение своими высокими гуманистическими принципами тенденции в создании и использовании открытых (бесплатных) образовательных ресурсов (Open Source) и программного обеспечения, которые обеспечивают поддержку мировой системы образования. Так, например, большое распространение среди Российских вузов получил программный комплекс с открытым кодом Moodle, который является специализированной системой управления учебным процессом – LMS с использованием сети Интернет. На сайте [www.Moodle.org](http://www.Moodle.org) находится документация, инсталляционные пакеты последней версии, а также средства онлайн-поддержки пользователей и разработчиков. Важнейшей особенностью Moodle является возможность сохранения курсов в стандартизованном виде, что позволяет проводить обмен курсами не только между факультетами, но и между вузами. Кроме того, учебный курс в среде Moodle может оперативно корректироваться преподавателем в соответствии с теми или иными целями, т.е. развиваться и поддерживать свою актуальность. По дидактическим возможностям Moodle не уступает многим платным программным средам, поэтому именно её выбираем в качестве инструментария для разработки виртуальной составляющей D-системы.

*Требование четвертое.* D-система должна формировать среду опережающей подготовки (опережающего обучения).

Опережающая подготовка предусматривает такую организацию обучения, при которой содержание профессионального образования и технология его реализации формируют у будущих специалистов зону ближайшего развития (Л.С. Выготский), которая позволит им не только адаптироваться и утвердиться в своей профессиональной деятельности, но и сформирует готовность к постоянному совершенствованию. Для этого в содержании образования акцент с достигнутого (актуального) уровня развития производства, науки и техники должен быть перенесен на перспективный (потенциальный), а педагогические технологии проектироваться с учетом постоянного обновления содержания.

Однако осуществление опережающего обучения, особенно его практической составляющей, сопряжено с такой трудностью как, непрерывно обновляющаяся номенклатура технических средств и инструментария,

которая требует постоянного совершенствования лабораторной базы и разработки методик ее использования в образовательном процессе.

Поделимся опытом в решении этого вопроса.

Прежде всего, мы провели анализ выпускаемого в России и странах СНГ лабораторного оборудования. Убедились в том, что для блока дисциплин: электроника, цифровые устройства и микропроцессоры, вычислительные средства в измерительной технике и т.п., готового решения, удовлетворяющего поставленной задаче опережающего обучения, нет. Пришлось искать его самостоятельно. С этой целью проанализировали содержание дисциплин и методы обучения и соотнесли их с потребностями современного производства, достижениями науки и техники. Оказалось, что для блока рассматриваемых дисциплин характерны: 1) динамично совершенствующаяся элементная база, 2) методы обучения, базирующиеся на информационных технологиях и 3) тесные междисциплинарные связи.

*Первое.* Содержание обучения по названным дисциплинам должно быть наполнено сведениями о современных достижениях науки и техники, применяющихся в промышленных, аграрных и транспортных технологиях. Логика познания (преподавания) требует включения в содержание дисциплин информации о принципах действия и характеристиках новых компонентов и функциональных узлов, а также об элементах программирования, участвующих в выработке нового технического мышления при управлении современными измерительными приборами и системами. Обновление элементной базы происходит настолько быстро, что производство лабораторных стендов – монолитов с общим корпусом, встроенными приборами и источниками питания – рассчитанных на подготовку поколения специалистов, – оказалось нецелесообразным. Поэтому на современном рынке учебного оборудования образовался пробел – дефицит лабораторных стендов по электронике, цифровым устройствам и микропроцессорам, вычислительным средствам в измерительной технике, укомплектованных новыми компонентами и построенных по иерархическому принципу кибернетических систем.

*Второе.* Система образования должна активно осваивать и адаптировать к процессу обучения информационные технологии. Так, например, различные среды схемотехнического моделирования успешно вошли в образовательную практику многих вузов. Они позволяют моделировать различные электромагнитные явления и процессы, визуализируют их и существенно ускоряют математическую обработку. Поэтому включение компьютерного моделирования в лабораторные практикумы, курсовые и дипломные работы стало необходимым фактором развития современной системы обучения. На сегодняшний день отечественной и зарубежной практикой инженерного образования освоены такие среды схемотехнического моделирования как Electronics Workbench, Multisim, Mathcad, MATLAB и др. Процесс освоения новых технологий будет постоянным: чтобы готовить специалистов для передового производства, система образования должна у него учиться.

*Третье.* Тесные междисциплинарные связи в блоке «электроника, цифровые устройства и микропроцессоры, вычислительные средства в измерительной технике и т.п.» позволяют комплексно решать в процессе обучения задачи формирования профессиональных компетенций.

Эти соображения дают лишь общие ориентиры для дидактического проектирования лабораторной базы опережающего обучения. Они объясняют отсутствие готового промышленного решения, что, кстати сказать, осознается не только преподавателями, но и производителями. Поэтому первый из предлагаемых проектов был выработан в ходе взаимовыгодного сотрудничества с производителями аппаратных средств.

Выбирая технические средства для переоснащения лабораторий, мы остановились на продукции научно-производственных объединений (НПО) «ОВЕН» (Москва) и «ЭНЕРГОМЕРА» (Ставрополь). Эти НПО имеют широкую дилерскую сеть на территории России и стран СНГ и успешно внедряют свои разработки за счет высокого качества и надежности, современного эстетического дизайна, а также доступности цен и гибкой системы скидок. Существенным преимуществом НПО «ОВЕН» и «ЭНЕРГОМЕРА» по сравнению с

другими производителями является наличие программы поддержки вузов (бесплатная поставка оборудования), что сделало наш выбор окончательным.

Рассмотрим и другой подход к созданию лабораторной базы опережающего обучения. Если в первом варианте использовались приборы и устройства НПО «ОВЕН» и «ЭНЕРГОМЕРА», а техническое и дидактическое проектирование, монтаж и наладка осуществлялись сотрудниками вуза, то во втором варианте в основу проектируемых лабораторных стендов легли запатентованные научные разработки преподавателей.

Известно, что в содержании блока дисциплин: электроника, цифровые устройства и микропроцессоры, вычислительные средства в измерительной технике и т.п. междисциплинарным учебным модулем является «микропроцессорная техника».

В последнее время широкое распространение в практике использования микропроцессорной техники получили микроконтроллеры. Область их применения многогранна и охватывает информационно-измерительные системы и системы управления технологическими процессами, т.е. системы, реализующие основные информационные функции: сбор, хранение, обработку, передачу и использование информации. К ним относятся интеллектуальные датчики различного назначения, спутниковые навигационные системы, автомобильные системы управления, медицинская техника, автоматизированные системы управления различными объектами и процессами. Знание структуры, функциональных возможностей микроконтроллеров, методов и средств разработки устройств и систем на их основе является необходимым для каждого специалиста в области электроники, вычислительной и измерительной техники.

Микроконтроллер является законченным устройством. Студенту остается выбрать наиболее подходящий микроконтроллер, подключить к нему, датчики, клавиатуру, индикатор, ключи и т.д., а также разработать программу. Наиболее сложная и трудоемкая часть микроконтроллерного устройства – это программа. Отсутствие учебно-методического обеспечения, ориентированного на подготовку студентов к программированию, существенно тормозит освоение микроконтроллеров. Это проявляется в том, что в большинстве студенческих дипломных работ и далее диссертационных исследований, затрагивающих области применения электроэнергии и содержащих разработки цифровых электронных устройств отсутствуют микроконтроллеры, что, существенно снижает качество этих работ. Поэтому первой задачей в создании лабораторной базы опережающего обучения была разработка учебного пособия «Введение в программирование микроконтроллера AVR на языке Ассемблера», оказывающего практическую помощь студентам и аспирантам в освоении процесса разработки программ для популярных микроконтроллеров семейства AVR, корпорации Atmel. Эти устройства по соотношению цена – производительность – энергопотребление занимают лидирующее место на мировом рынке микроконтроллеров.

В пособии [Вострухин, 2010] рассмотрены вопросы программирования типовых функций микропроцессорных систем управления (сбор, хранение, обработка, передача и использование информации), разработана система заданий для их отработки, приведены примеры выполнения каждого из видов заданий и разработаны тесты контроля и самоконтроля.

Для отработки практических навыков программирования создан стенд микроконтроллерный, на который получено положительное решение о выдаче патента на промышленный образец. Стенд состоит из двух модулей, имеющих одинаковый набор основных компонентов рис. 2, но отличаются друг от друга принципиальными схемами.

Модули соединяются между собой проводниками. Варианты соединения модулей могут быть различными и зависят от того, какая задача стоит перед студентами. Например, модуль 1 можно использовать для преобразования аналоговых сигналов в частоту или ШИМ сигналы, которые можно подавать на вход модуля 2 для дальнейшей их обработки и вывода на индикатор. Таким образом, оба модуля представляют единое целое. Это расширяет функциональные возможности учебного стенда.



Так как модуль 1 позволяет разрабатывать АЦП циклического, следящего и последовательных приближений, то на базе стенда микроконтроллерного можно создавать цифровые вольтметры и автоматические регуляторы, работающие от датчиков с аналоговым выходом в форме унифицированных сигналов напряжения или тока, а также от термопреобразователей сопротивления типа ТСМ или ТСР. На основе модуля 2 можно разрабатывать: частотомеры, измерители временных параметров ШИМ сигналов, генераторы прямоугольных импульсов, ряд измерительных и преобразовательных устройств.



Рис. 2. Структура модуля стенда микроконтроллерного

## Заключение

Разработанный образовательный ресурс учебного модуля «Микропроцессорная техника» представляет собой совокупность содержания, изложенного в учебном пособии в форме инновационной практико-ориентированной задачи с квалифицированным дидактическим сопровождением этапов ее решения и инструментария – стенда микроконтроллерного (рис. 3).



Рис. 3. Образовательный ресурс учебного модуля «Микропроцессорная техника»

В процессе апробации и внедрения стенда микроконтроллерного в учебный процесс на его основе реализованы: система управления однофазным мостовым инвертором напряжения; АЦП циклического и следящего типов с программно управляемой разрешающей способностью; двухпозиционный терморегулятор с аналоговым входом; генератор прямоугольных импульсов; ряд измерительных и управляющих устройств.

Правильность выбранного решения подтверждается примерами современной образовательной практики в области освоения микропроцессорной техники (МИФИ – лаборатория «Микропроцессорные системы»; Ивановский государственный энергетический университет – авторы М.Н. Горячев и А.В. Волков;

Харьковский государственный автомобильно-дорожный технический университет – авторы Л.С. Абрамов и А.А. Буты; Белорусское научно-техническое предприятие «Центр» и др.).

В заключении подчеркнем, что образовательные ресурсы с широким спектром дидактических и эстетических свойств, направленных на формирование компетенций в области создания и применения микропроцессорной техники востребованы в современном инженерном образовании. Дидактический дизайн становится научно-методической платформой разработки таких ресурсов. Многообразие отечественных и зарубежных научных школ предполагает разнообразие и уникальность решений в создании образовательных ресурсов. Соответствие дидактических возможностей требованиям опережающей подготовки инженеров, оптимальность соотношения цены и качества определит лучшие среди них.

---

## Библиография

- [Похолков, 2000] Похолков, Ю.П. Основные принципы национальной доктрины инженерного образования / Ю.П. Похолков, Б.Л. Агранович // Томский политехнический университет, 2000.
- [Горохов, 2007] Горохов, В. Инновационно-ориентированная оценка техники как вид социотехнического проектирования / В. Горохов // Alma mater. 2007. № 9. С. 18-24.
- [Соснин, 2009] Соснин, Н.В. Дизайн как основа компетентностной модели инженерного образования / Н.В. Соснин // Высшее образование в России. – 2009. – № 12. – С. 24.
- [Нуриев, 2009] Нуриев, Н.К. Проектирование дидактических систем нового поколения для подготовки способных к инноватике инженеров / Н.К. Нуриев, Л.Н. Журбенко, С.Д. Старыгина, Е.В. Пашукова, А.Р. Ахмадеева // Образовательные технологии и общество. – 2009. – №12. – С. 417-440.
- [Бутырин, 2009] Бутырин, П.А. Развитие высшего электротехнического образования в России / П.А. Бутырин // Электричество. – 2009. – № 8 С. 7-11.
- [Вострухин, 2010] Вострухин, А.В. Введение в программирование микроконтроллера AVR на языке Ассемблера: учебное пособие / А.В. Вострухин, Е.А. Вахтина. – 2-е изд., перераб. и доп. – М.: Илекса, 2010. – 184 с.

---

## Сведения об авторах



**Вахтина Елена Артуровна** – Ставропольский государственный аграрный университет, кандидат педагогических наук, доцент кафедры «Автоматика, электроника и метрология»; Россия, 355017, Ставрополь, пер. Зоотехнический, 12; e-mail: [VEA1961@yandex.ru](mailto:VEA1961@yandex.ru)

Главные области научного исследования: инженерная педагогика, теория и методика e-learning



**Вострухин Александр Витальевич** – Южно-Российский государственный университет экономики и сервиса, Ставропольский технологический институт, кандидат технических наук, доцент кафедры «Информационные технологии и электроника»; Россия, 355035, Ставрополь, ул. Кулакова, 41/1; e-mail: [avostrukhin@yandex.ru](mailto:avostrukhin@yandex.ru)

Главные области научного исследования: микроконтроллерные измерительные преобразователи физических величин

## ОБРАЗОВАТЕЛЬНЫЕ ПОРТАЛЫ ДЛЯ ШКОЛ – ОЖИДАНИЯ И РЕАЛЬНОСТЬ

Диана Богданова

**Abstract:** Приведены результаты анализа работы поисковой системы образовательного портала школ Евросоюза – Learning Resource Exchange. Приведены возможные способы классификации цифровых образовательных ресурсов по типу и по содержательному наполнению. Описан процесс формирования Портала, проходивший в несколько этапов. Приведены критерии, положенные в основу оценки рассматриваемых ресурсов. Описана процедура оценивания ресурсов для их размещения в Learning Resource Exchange. Рассмотрены и проанализированы отдельные примеры цифровых образовательных ресурсов, представленных в портале, с учетом опыта работы автора с образовательными ресурсами российских порталов. Анализ проводился на основе критериев качества ресурсов, разработанных и предложенных в Learning Resource Exchange. Рассмотрены метаданные отдельных ресурсов. Предложен путь совершенствования работы поисковой системы. Предложен способ совершенствования системы контроля качества ресурсов, размещаемых в Learning Resource Exchange.

**Keywords:** Образовательный портал, поисковая система, цифровой образовательный ресурс, метаданные, критерии качества цифровых образовательных ресурсов

**ACM Classification Keywords:** K.3.1 [Computer Uses in Education]: Computer-assisted instruction (CAI), Distance learning

---

### Введение

В последнее десятилетие в процессе информатизации школьного образования, как в России так и за рубежом, произошел стремительный рост количества цифровых образовательных ресурсов и образовательных сетей, порталов, содержащих эти ресурсы. Так, 2008 году начал работать в полноценном режиме российский сайт Единой коллекции цифровых образовательных ресурсов ([www.school-collection.edu.ru](http://www.school-collection.edu.ru)). При этом ресурсы, представленные в Единой коллекции, можно условно разделить на три типа [Осин, 2008]. Первый — текстографические ресурсы, то есть, тексты, иллюстрированные некоторыми картинками. По существу это – цифровая копия учебника, которая может храниться в электронных библиотеках. Второй — монотипные ресурсы, к которым можно отнести аудиозаписи, видеофрагменты, картины - то, что выполнено в одном формате. И, наконец, третий тип— мультимедийные образовательные ресурсы. Причем, ресурсы первого и второго типов составляют в Единой коллекции большинство. По содержательному наполнению ЦОР можно классифицировать по нескольким категориям: поддерживающие существующие традиционные учебники; предлагающие информационные ресурсы сложной структуры; предлагающие инновационные учебно-методические комплекты [Христочевский, 2008]. К моменту запуска сайт Единой коллекции содержал порядка 85 000 цифровых образовательных ресурсов, и их число до сих пор росло и продолжает расти.

В процессе использования содержимого Единой коллекции обнаружилось недоработки как в содержании самих ресурсов, так и в работе поисковой системы, что отчасти можно было объяснить не всегда корректной подготовкой метаданных. В связи с этим возникла проблема устранения выявленных недоработок, корректировки ресурсов и совершенствования системы поиска. Эти работы велись одновременно с пополнением содержимого Коллекции и совершенствованием критериев и методов отбора ресурсов.

Практически одновременно с Россией аналогичные работы по созданию образовательных порталов велись и за рубежом. Для российских специалистов особый интерес представлял проект, разрабатывавшийся для школ Европейского союза, а именно, его часть, связанная с цифровыми образовательными ресурсами - Learning Resource exchange (LRE) - Обмен Образовательными Ресурсами, открытый для использования в третьем квартале 2009 года.

Анализу ресурсов и работы навигационной системы LRE и посвящается эта работа.

---

### Описание объекта исследования

---

Обширное содержимое портала доступно по лицензии Creative commons. Его разработка велась планомерно, поэтапно, на протяжении более 6 лет, в его основу легли проекты ASPECT, CALIBRATE и MELT. Проект ASPECT ставил целью разработку стандартов нахождение нужного образовательного содержания и его использование. 40 школ в 4 странах использовали настроенную под них версию LRE для того, чтобы определить, каким образом использование предлагаемых стандартов и спецификаций способствует наилучшему использованию содержимого LRE. Проект CALIBRATE завершился в середине 2008 года. В нем участвовали 8 стран, в основном новых членов Евросоюза. В его задачу входило оказание помощи Министерствам образования этих стран в получении доступа к хранилищам образовательных ресурсов, а также поиск новых путей к использованию образовательных материалов в европейском образовательном web – сообществе. И, наконец, проект MELT, основной задачей которого являлось обеспечение возможности использования образовательного содержимого не только в пределах одной страны, опираясь на различные педагогические подходы, но также и предоставление возможности учителям европейских стран находить и по-своему использовать образовательные ресурсы и элементы из других стран Евросоюза. Проект MELT предложил новые пути для экономичного расширения содержимого образовательных ресурсов. В странах Евросоюза результатом проекта MELT стало увеличение «критической массы» существующих образовательных материалов, а также разработанные принципы оценки качества ресурсов.

Базовый принцип, положенный в основу оценки качества, состоял в том, чтобы, все предлагаемые к доступу образовательные материалы, первоначально проходили проверку на национальном уровне. Эта проверка должна была гарантировать правильность и точность содержания, отсутствие материалов, носящих оскорбительный характер или содержащих угрозы, а также, что авторское право не было нарушено. Известен ли автор ресурса или нет, был ли ресурс разработан учителем или коммерческими компаниями - он становился доступным через только после процедуры первоначального рассмотрения на национальном уровне. При этом в состав национального экспертного совета могли быть приглашены опытные учителя-предметники, а также специалисты в предметной области. Им отводилась роль модераторов.

Проверка содержимого ресурса проводилась, как правило, по пяти направлениям: педагогическое, удобство использования, возможность повторного использования, доступность и техническая сторона. Список категорий имел рекомендательный характер, т.к. не все критерии могли быть одинаково успешно применены ко всем ресурсам. Так, например, некоторые ресурсы имеют хорошую структуру и могут быть легко адаптированы к обучающим сценариям, отличным от тех, что изначально планировался разработчиком. Однако эти же ресурсы, например, могут не отвечать требованиям интерактивности. Таким образом, проверочный список рассматривался как минимальный набор, который можно использовать очень гибко при проверке содержимого ресурса. Помимо этого, были разработаны критерии качества для интернационального использования ресурсов. Такие ресурсы получили статус «travel well» - «хорошо путешествующих». С точки зрения здравого смысла можно предположить, что отдельные ресурсы будут «путешествовать» лучше, чем другие, так как используют фотографии, графику, звук. Но, помимо этого, вероятность получить статус «хорошо путешествующих» выше у тех обучающих единиц, структура которых:

- A) модульна и позволяет их трансформировать в случае необходимости,
- B) имеет хорошую визуальную составляющую,
- C) либо двуязычна, либо позволяет легко перевести с языка оригинала на другой язык,
- D) тема ресурса должна быть транснациональна с точки зрения учебных программ и,
- E) безусловно, должна иметь лицензию Creative Commons.

Ознакомившись с работой поисковой системы, автор приступил к сёрфингу по portalу. Поиск интересующего ресурса можно осуществлять тремя возможными путями:

- по ключевому слову;
- по предмету, выбирая его из списка предметов, при необходимости используя фильтр «язык» и «возраст учащихся»;
- по тэговому облаку.

Основными критериями выбора ресурсов для рассмотрения был «язык ресурса – это должен быть английский, либо отсутствие языка – language free», а предметы выбирались из тех, в которых автор считал себя достаточно компетентным.

### Анализ ресурсов раздела ИКТ

Первым были рассмотрены ресурсы по ИКТ. В списке найденных первым привлек внимание ресурс, содержательно сфокусированный на историю Древнего Рима, содержащий вопросы, такие, например, как «Носили ли древние римляне тоги?». Ключевыми словами к этому ресурсу даны «ИКТ» и «рабочий материал», хотя по содержанию он кажется ближе к предмету истории. Как оказалось, этот ресурс создан в поддержку компьютерной игры на CD-дисках «Ancient Lands», посвященной Египту, Риму и Греции, разработки Microsoft 1994 года. Хочется поприветствовать игру-долгожителя. Игра существует на CD-диске, а ресурс в поддержку к ней размещается в образовательном portalе.

В этом же списке ресурсов по ИКТ, только не рядом с упомянутым выше ресурсом, а чуть дальше, по неясной причине оказалась и разработка Массачусетского технологического института (МТИ) по психологии.

Привлекли внимание имеющие статус «Language free» ресурсы по ИКТ. При более близком знакомстве ими оказались несколько вариантов фотографий компьютерных мышей, клавиатуры, а также фото платы, адаптера и целиком компьютера с процессором и монитором. (Рис. 1-3).

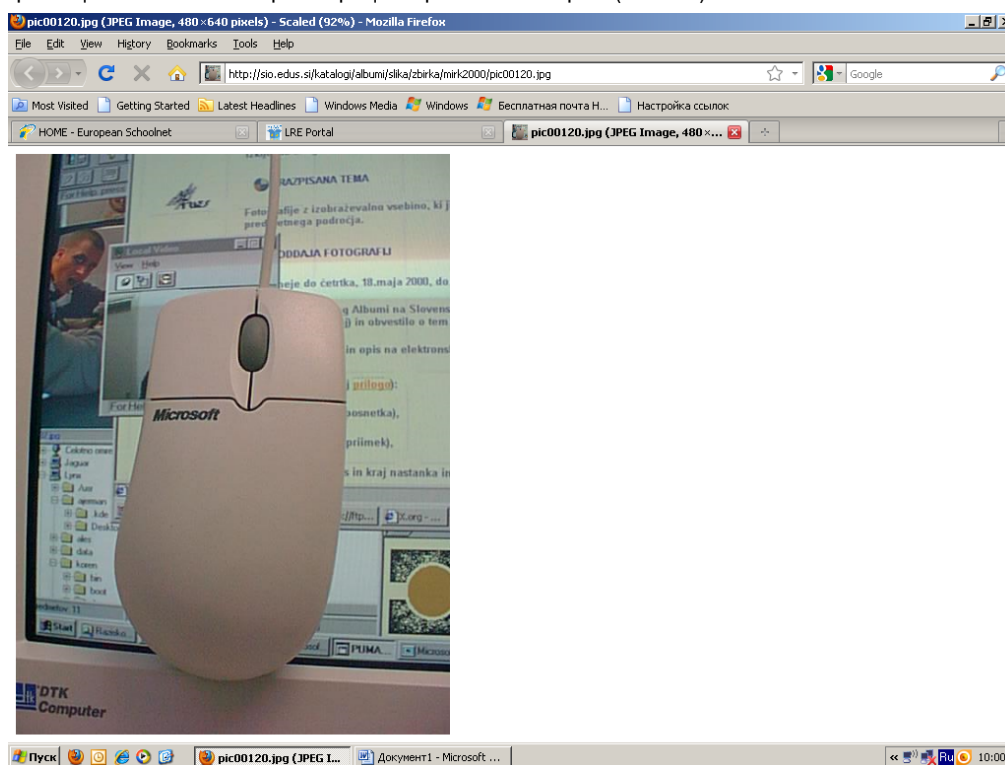


Рисунок 1.

Причем, в метаданных к приведенному выше ресурсу сказано – мышь, как типичное устройство ввода.

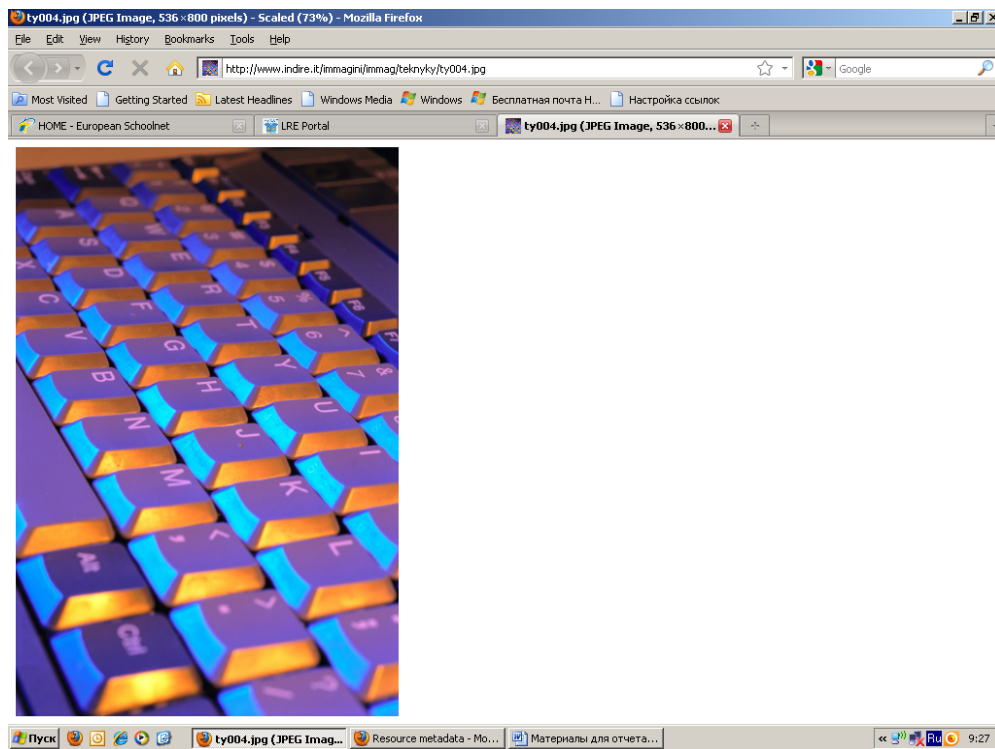


Рисунок 2.

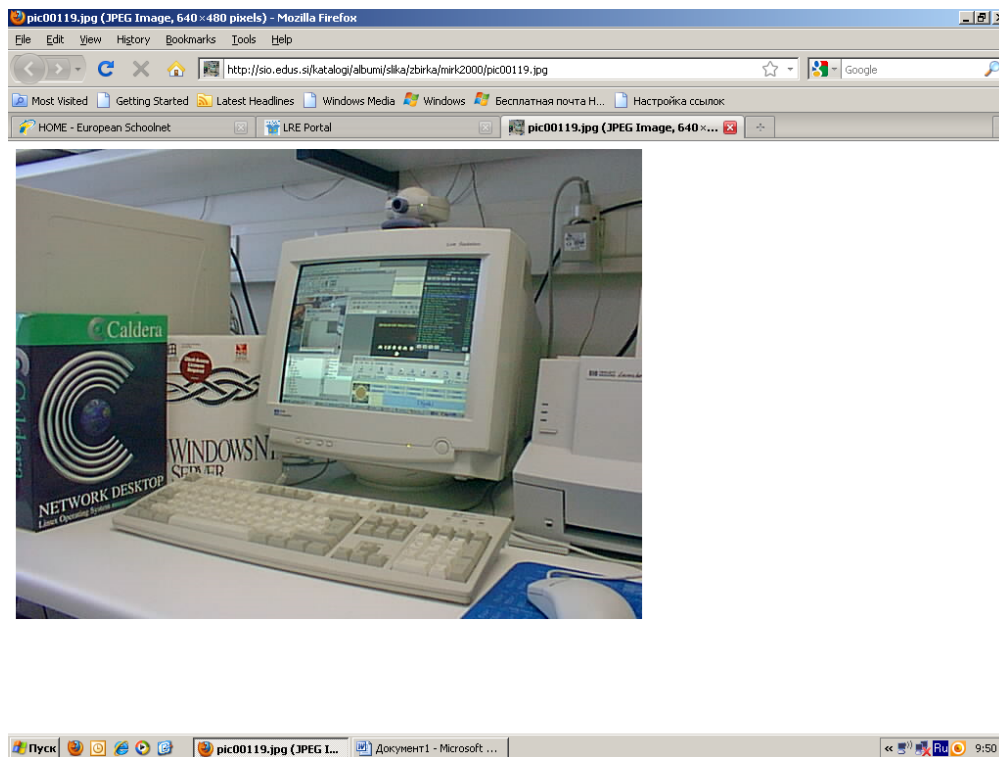


Рисунок 3.

Там же, в разделе ИКТ оказались 4 фотографии малыша, видимо сына автора ресурса, в разное время года с названием «Миха 4 сезона». Ресурсы содержали только фотографии без каких-либо комментариев, имели ключевые слова: фотография, время года, обработка цифрового фото. Осталось неясным, каково дидактическое наполнение этих ресурсов.

Увиденное не могло не вызвать изумления. Каким же образом эти ресурсы проходили отбор? Если руководствоваться перечисленными выше критериями качества для отбора ресурсов, то:

- правильность и точность содержания не нарушена. В метаданных этих ресурсов указано «рабочая станция», «мышь» и т.д.;
- дидактическое содержимое ресурса оценить невозможно;
- удобство использования - соответствие этому требованию вызывает сомнение и улыбку. Ситуация парадоксальная: фото компьютера и его компонентов размещено в качестве цифрового образовательного ресурса.
- по критериям отбора, с точки зрения возможности повторного использования, доступности и технической стороны они вполне удовлетворяют объявленным требованиям.

А по классификации LRE эти ресурсы могли бы получить статус «хорошо путешествующих».

Показался очень полезным ресурс, представляющий курс Открытого Университета, посвященный технологии поиска информации в Интернете. Он учит, как выбирать ключевые слова, навигационные системы, предметные шлюзы, он-лайн базы данных и т.д. Этот курс современный, своевременный и очень актуальный.

---

### **Анализ ресурсов раздела Технологии**

---

Далее были рассмотрены ресурсы по предмету «Технологии». Здесь также в списке ресурсов среди других опять оказались разработки МТИ, на этот раз – это был курс китайского языка, а также история евреев с библейских времен до наших дней, ресурс по нейрофизиологии, по финансовому рынку и законодательству в корпоративных финансах. Могло бы возникнуть предположение, что в случае поиска по предмету «Технологии» поиск осуществляется не по предмету, а по «Технологии», которая в данном случае является ключевым словом и содержится в названии разработчика, и что результаты поиска, возможно, имеет смысл перепроверять по метаданным. Однако другие погрешности работы поисковой системы, упомянутые выше, этим предположением не объяснить. Эти ресурсы, как и подавляющее большинство ресурсов МТИ, а также представленных разработок Открытого Университета (Open University) имели возрастную рекомендацию 18+. Но, насколько известно автору школьное обучение в европейских школах заканчивается к 18 годам. Следует отметить, что все разработки МТИ, что были просмотрены автором, хотя и вызывали сомнение правомерностью своего пребывания в школьных ресурсах, выполнены на хорошем уровне.

---

### **Анализ ресурсов раздела Хорошо путешествующие**

---

Следующим предметом для рассмотрения была выбран ресурс из раздела «хорошо путешествующих» по теме логарифмы и логарифмическая функция (рис.4-5). Из метаданных следует, что представленная разработка доступна на английском, французском, португальском и еще на четырех языках. Описание теории логарифмов и логарифмической функции на заявленных языках представлено. Причем, в рекомендациях сказано, что прежде, чем переходить к построению 3-d функций, надо хорошо изучить построение в 2d. Когда же дело дошло построения 2-х и 3-х мерных графиков логарифмической функции, до чтения инструкции по установке необходимого для этого программного обеспечения и работы с программой, материал уже предлагался только на испанском языке. Но для того, чтобы прочитать теоретический материал про логарифмы, совершенно не обязательно обращаться в Интернет, достаточно просто открыть учебник математики на своем родном языке. А именно та часть ресурса, которая использует возможности компьютера для визуализации логарифмических функций и

интерактивной работы с ними, оказывается недоступной для тех, чьи языки были объявлены в списке ресурса.

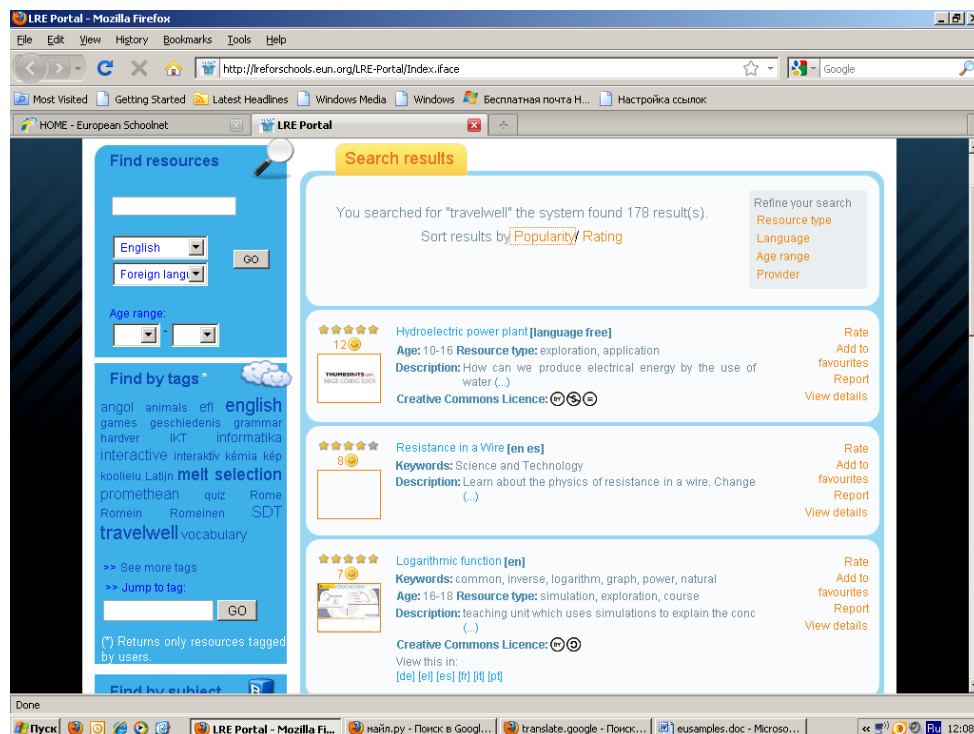


Рисунок 4.

Таким образом, при проверке содержимого по разработанным критериям для «хорошо путешествующих» ресурсов получается следующее: структура модульна; возможно, имеет визуальную составляющую, но проверить этого не удалось по причине незнания автором испанского языка; тема ресурса транснациональна; имеет лицензию Creative Commons. Это хорошая разработка для испанских школ, которая не должна была бы рекомендоваться для школ Евросоюза в своем нынешнем виде.

### Анализ ресурсов раздела Английский язык

Анализируя ресурсы по английскому языку, размещенные в российской Единой коллекции цифровых образовательных ресурсов, автор неоднократно отмечал в качестве недостатка прямой перенос материала из учебника на экран компьютера. В материалах меняется только тип носителя, при этом не используются мультимедийные интерактивные возможности компьютера, а от учителя не требуется изменения методики преподавания, т.е. речь идет о ресурсах первого типа

В разделе «Английский язык» LRE оказалось много ресурсов подобного рода. Так, например, некоторые немецкие ресурсы по английскому языку представляют собой просто отсканированные страницы учебника (Рис.6). И в этом случае возникает недоумение, каким же образом эти ресурсы проходили проверку по критериям, упомянутым выше. Среди просмотренных из представленных ресурсов по языку автору показался интересным многоязычный интерактивный репетитор по глаголам. По своей идее он соответствует философии Евросоюза.



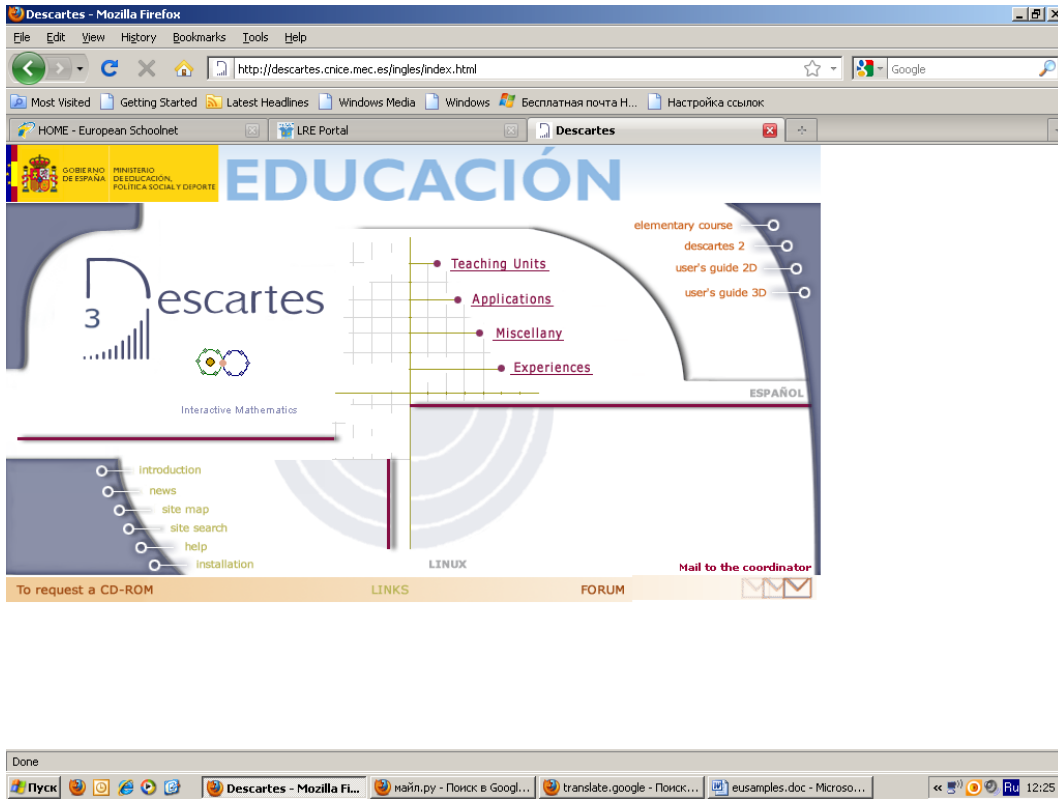


Рисунок 5.

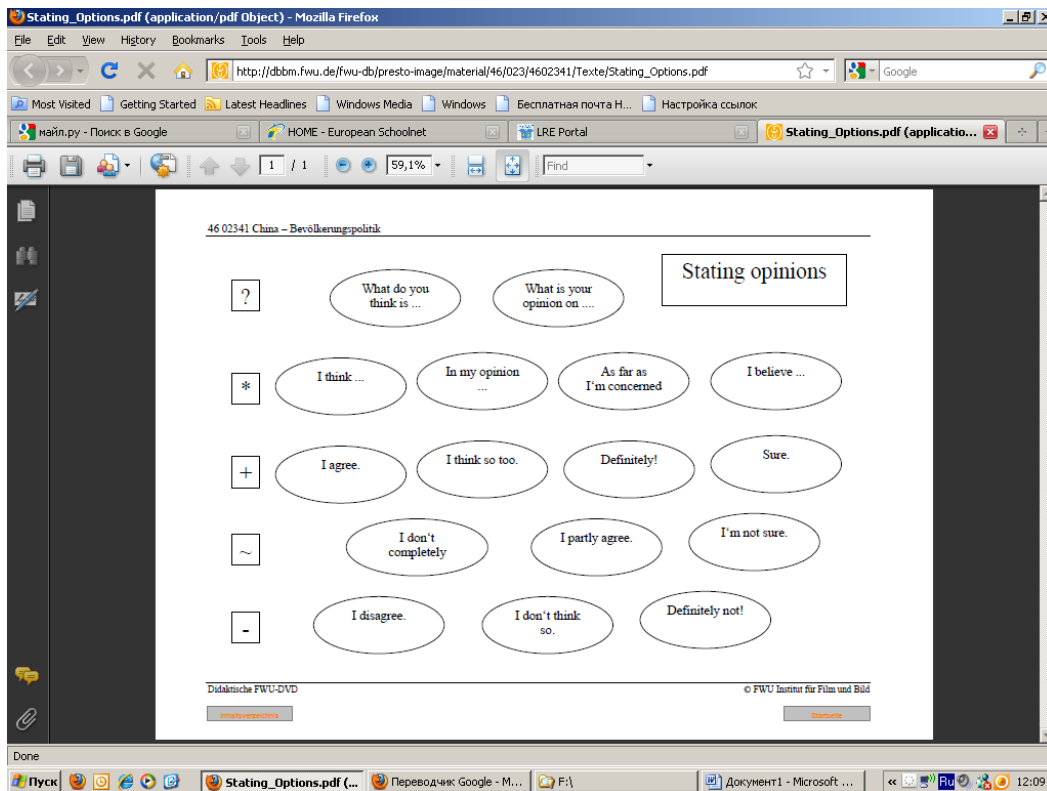


Рисунок 6.

### Заключение

---

Можно бесконечно путешествовать среди ресурсов, представленных в LRE. И это путешествие будет очень интересным, несмотря на попадания в тупик, выходы в неожиданные участки, не соответствующие первоначальному описаниям. Опираясь на опыт работы с Российскими образовательными порталами и изучение LRE Евросоюза, можно сделать вывод о том, что создание образовательных порталов все еще проходит стадию становления. Ни в России, ни в Евросоюзе пока не удалось создать безошибочно работающей поисковой системы, что предъявляет дополнительные требования к организации метаданных ресурсов. Хотя, безусловно, помимо требований к аккуратному заполнению метаданных, следует совершенствовать работу и самих поисковых систем.

Сами же ресурсы, во всяком случае, значительное их количество, требует серьезной проверки на соответствие требованиям, к ним предъявляемым. Требования могут варьироваться для конкретных случаев, но эффективное использование возможностей компьютера должно быть основополагающим. Не должны и не могут существовать цифровые образовательные ресурсы, содержащие фотографии компьютерных мышей, клавиатуры и т.д. Они просто не должны называться таковыми.

Возможно, что для проверки соответствия объявленным требованиям имеет смысл иметь отдельную экспертную комиссию, а имя эксперта, подтвердившего соответствие требованиям, включать в список метаданных ресурса вместе с именами разработчиков.

---

### Bibliography

---

[Осин 2008 ] А.Осин. Интернет-конференция “Школьная информатизация: образовательные ресурсы, программное обеспечение, Интернет”. 28.04.2008 12:49 генеральный директор Федерального государственного научного учреждения «Республиканский мультимедиа центр».

[Христочевский, 2008] С.А. Христочевский «Проверено, Рекомендовано. Единая коллекция ЦОР и пути ее совершенствования». «ИКТ в образовании»19.2008.14.10.2008

European Schoolnet. Transforming education in Europe. <http://reforschools.eun.org/LRE-Portal/Index.iface>

---

### Author's information

---



**Diana Bogdanova** – Science researcher, Ph.D., The Institute of Informatics Problems of the Russian Academy of Sciences (IPI RAN) [www.ipiran.ru](http://www.ipiran.ru) e-mail: [d.a.bogdanova@mail.ru](mailto:d.a.bogdanova@mail.ru)

Major Fields of Scientific Research: ICT for school education, Digital educational resources, Safe Internet, Cyber threat, Social networks, Media literacy

## ЭЛЕМЕНТЫ АДАПТИВНОЙ ТЕХНОЛОГИИ ОЦЕНИВАНИЯ ЗНАНИЙ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ

Константин Юрченко, Виталий Снитюк

**Аннотация:** Рассмотрены особенности реализации компьютерных систем профессиональной подготовки с учетом их адаптации в режиме реального времени. Разработаны процедуры коррекции сложности вопросов в зависимости от уровня знаний каждого обучаемого. Предложена эвристика о сходимости сложности вопроса к постоянному значению.

**Ключевые слова:** компьютерные системы профессиональной подготовки, сложность, адаптация.

**ACM Classification Keywords:** I.2.1 - Applications and Expert Systems.

---

### Введение

Последнее столетие и, особенно, его последние десятилетия отмечены качественным рывком в разработке и использовании информационных технологий. Такой вывод базируется на трех «китах»: создании вычислительной техники, разработке соответствующего программного обеспечения и применении первых двух составляющих к решению все более сложных прикладных задач. Развитие глобальной сети и телекоммуникаций – еще один аспект, определяющий современное состояние производственной сферы и инфраструктуры. Широкое проникновение информационных технологий во все области жизнедеятельности человека привело к тому, что значительная часть мировых финансовых потоков генерируется и проходит через информационно-коммуникационную сферу. В связи с этим возникло обоснованное предложение считать современный этап развития «информационным обществом». С другой стороны, экспоненциальный рост информации в мире и ее распространение через сеть Интернет, а также развитие самой сети и ее новые сервисы дают основания называть этот этап «обществом без границ». Свободное распространение информации, рост ее хранилищ, извлечение новых знаний и использование для этих целей территориально удаленных кластерных и Grid-систем, указывает на формирование «общества, базирующегося на знаниях» [Згуровский, 2005].

Все указанные выше аспекты, являются причиной и основанием для разработки экспертных и обучающих систем, в которых интегрированы возможности глобальной сети, удаленного обучения и получения новой информации. Известно, что первые экспертные системы были разработаны в середине 20-го столетия, а промышленные применения нашли в 80-х годах. Сегодня сложно представить вид деятельности, где не была бы задействована вычислительная техника и где бы не делались попытки применения экспертных систем.

Одной из таких областей является образование. Первые шаги в создании экспертных систем в сфере образования сводились к разработке систем тестового контроля знаний учащихся [Глибовец, 2000]. И сегодня почти в каждом высшем учебном заведении Украины используют одну, а то и несколько таких систем автоматизированного контроля знаний. На следующем этапе предлагались системы обучения, которые с развитием мультимедийных технологий вывели на новый уровень преподавание отдельных дисциплин. Интеграция систем обучения и контроля позволила замкнуть цикл обучения. Жесткость такой структуры стала причиной внедрения в автоматизированные системы адаптационных механизмов, базирующихся на различных принципах и моделях [Коджа, 2003].

Авторами ранее уже были выполнены некоторые шаги в направлении формирования элементов методологии разработки автоматизированных систем обучения и контроля знаний (АСОКЗ) и проведения процессов обучения и контроля знаний. В частности, в работе [Снитюк, 2000] в основу формирования базы данных АСОКЗ предложено положить классификацию вопросов в зависимости от типов ответов: «Да-Нет»; один из многих; несколько из многих; число; интервал; нечеткий интервал; слово; предложение. Для каждого из этих типов вопросов разработаны процедуры оценки, исходя из отклонений ответов

обучаемого лица от правильных ответов и метод формирования интегральной оценки [Снитюк, 2003]. В следующей статье [Снитюк, 2008] было предложено осуществлять структуризацию предметной области, которой, чаще всего, и есть учебные курсы, что позволило оптимизировать процессы тестирования за счет самоорганизации множества вопросов, которые задаются обучаемому. Аспекты создания эффективных компьютерных систем профессиональной подготовки как одного из видов АСОКЗ рассмотрено в [Снитюк, 2010], где определено, что такие системы должны удовлетворять требованиям: минимизации времени обучения и контроля знаний; полноты контроля знаний; отсутствия или минимального присутствия информационной избыточности и информационной недостаточности, максимально возможной объективизации результатов оценивания.

Современные мировые тенденции нацелены на создание таких АСОКЗ, которые ориентированы на определенную модель пользователя. В частности, создаются системы, учитывающее психологическое состояние обучаемого и соответствующим образом на него реагирующие. Не достаточно изученной является проблема объективизации уровня сложности задаваемых вопросов и, соответственно, определения интегральной оценки знаний.

### Адаптация уровня сложности элементов процесса контроля знаний

Предложим один из путей адаптации уровня сложности вопросов к оцениваемому множеству обучаемых. При этом будем базироваться на нескольких эвристиках:

Эвристика 1. Если обучаемый дает правильный ответ, то сложность вопроса уменьшается и наоборот.

Эвристика 2. Динамика значения уровня сложности зависит от количества обучаемых, знания которых контролируются, а также от суммарной оценки каждого из них.

Эвристика 3. Если обучаемый со сравнительно высокой суммарной оценкой дает правильный ответ, то сложность вопроса уменьшается на меньшее значение, чем в случае, когда правильный ответ дает обучаемый с низшей суммарной оценкой.

Эвристика 4. Если обучаемый со сравнительно высокой суммарной оценкой дает неправильный ответ, то сложность вопроса увеличивается на большее значение по сравнению со случаем, когда неправильный ответ дает обучаемый с низшей суммарной оценкой.

Эвристика 5. Минимальная сложность вопроса должна быть больше нуля.

Эвристика 6. Максимальная сложность вопроса должна быть меньше единицы.

Предположим, что оцениваются знания  $m$  обучаемых, количество вопросов -  $n$ . Коррекция сложности вопросов производится по формуле

$$p_i^j = p_i^{j-1} + f_i(p_i^0, m, d^{j-1}, Z), \quad (1)$$

где  $p_i^j$  - сложность  $i$ -го вопроса при контроле знаний  $j$ -го обучаемого,  $f_i$  - некоторая функция, в соответствии со значением которой осуществляется коррекция сложности  $i$ -го вопроса,  $p_i^0$  - начальная сложность вопроса,  $d^{j-1}$  - суммарная оценка ответов  $(j-1)$ -го обучаемого,  $d^{j-1} \in [0, 1]$ ,  $Z$  - некоторая величина, зависящая от величины отклонения данного ответа от правильного,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ . Как следует из формулы (1), коррекция уровня сложности вопроса производится сразу же после контроля знаний каждого обучаемого, т.е. в режиме реального времени.

Рассмотрим отдельные варианты формулы (1) в применении к вопросам различного типа.

Вопросы с ответами типа «Да-Нет». В таком случае коррекция сложности вопроса осуществляется по формуле

$$p_i^j = p_i^{j-1} \left( 1 + \frac{1}{m} (d^{j-1} \chi(i \text{ âĭ } \delta \text{ ââĕĕĭĭ } \hat{u} \acute{e} \hat{i} \text{ ò } \hat{a}\hat{a}\hat{o}) - (1 - d^{j-1}) \chi(i \text{ ðââĕĕĭĭ } \hat{u} \acute{e} \hat{i} \text{ ò } \hat{a}\hat{a}\hat{o})) \right), \quad (2)$$

где  $\chi(A) = \begin{cases} 1, & \text{если } A \text{ удовлетворяет } (1)-(6), \\ 0, & \text{иначе.} \end{cases}$  Очевидно, что для формулы (2) выполнены эвристики (1)-(6).

К другому типу относятся вопросы, где один ответ выбирается из нескольких, и правильность каждого определяется баллом (для  $i$ -го вопроса вектор баллов  $v = (v_1, v_2, \dots, v_k)$ , причем  $0 \leq v_i < 1 \forall i = \overline{1, k}$ ). Тогда коррекцию сложности вопроса в первом приближении осуществляем по формуле

$$p_i^j = p_i^{j-1} \left( 1 + \frac{1}{m} (d^{j-1} (\max_{l=1, k} v_l^{j-1} - a_i^{j-1}) - (1 - d^{j-1}) (a_i^{j-1} - \min_{l=1, k} v_l^{j-1})) \right), \quad (3)$$

где  $a_i^{j-1}$  - балл за ответ  $(j-1)$ -го обучаемого на  $i$ -й вопрос. Формула (3) открыта для внесения уточняющих коэффициентов и может быть использована для предварительных расчетов.

Адаптация сложности вопросов других типов осуществляется по аналогии с (1)-(3). Предположим, что все формулы для коррекции уровня сложности вопросов известны. Возможны две ситуации: начальные значения сложности вопросов известны и не известны. Как поступать в том и другом случаях?

Пусть начальные значения коэффициентов сложности вопросов заданы лицом, принимающим решение, экзаменатором или экспертом. Тогда в процессе оценивания знаний, реализуя случайный порядок экзаменуемых, получим колебания значений уровня сложности.

Эвристика 7. Предполагая бесконечное количество экзаменуемых, делаем вывод о сходимости значения уровня сложности вопроса к постоянному значению.

Заметим, что в формулах (2)-(3) предполагается учет влияния начального значения уровня сложности на весь дальнейший процесс его корректировки. Безусловно, от этого предположения можно отказаться, но только в том случае, если лицо, принимающее решение, дав начальные установки, стремится уклониться от дальнейшего активного участия в процессе оценивания. Выбор того или иного варианта зависит от экзаменатора и внешних обстоятельств.

Возможна ситуация, когда экзаменатор не может определить сложность вопросов с достаточной точностью. В таком случае начинать контроль знаний не следует, поскольку первые экзаменуемые получат значительно смещенные оценки. На помощь здесь может прийти моделирование. Экзаменатор, скорее всего, знает распределение учащихся, исходя из их учебных успехов (среднего балла в зачетной книжке). Если это не так, то можно воспользоваться различными дополнительными эвристиками, одна из которых указывает на то, что количество отличников составляет 20% от общего количества учащихся, количество учащихся на «хорошо» - 20% от оставшихся и т.д. В соответствие с распределением учеников, разыгрываем случайные значения, которые отвечают правильным или неправильным ответам, и осуществляем коррекцию значений сложности ответов. Такой подход приводит к объективизации процесса оценивания как в целом, так и его начального этапа в особенности, когда значения сложности вопросов носят сильно субъективизированный характер.

Если значения сложности вопросов не известны, то достаточно положить  $p_i^0 = 0,5$ ,  $i = \overline{1, n}$  и выполнить рассмотренные выше шаги.

---

## Заключение

В статье рассмотрен один из аспектов создания автоматизированных систем обучения и контроля знаний, а также компьютерных систем профессиональной подготовки. Отмечен низкий методологический уровень создания таких систем и предложен один из элементов реализации системного подхода к осуществлению дистанционного образования, а именно, к разработке подсистемы контроля знаний учащихся, базирующийся на адаптационных механизмах.

Сформулированы эвристики, которые положены в основу процедур коррекции уровней сложности вопросов в зависимости от вариантов ответов экзаменуемых и их общего образовательного уровня. Разработаны модели, с помощью которых осуществляется изменение сложности вопросов в режиме

реального времени, что позволяет уменьшить время обучения системы, а также минимизировать ошибки оценивания знаний.

Предложенный подход, равно как и другие существующие, требует реализации значительного объема верификационных процедур. Вместе с тем, он является достаточно простым для алгоритмизации и открытым для внесения изменений и дополнений.

---

## Литература

- [Згуровский, 2005] М.З. Згуровский, Н.Д. Панкратова. Системный анализ. Проблемы, методология, приложения. – К.: Наук. думка, 2005. – 743 с.
- [Глибовец, 2002] Н.Н. Глибовец, А.В. Олецкий. Искусственный интеллект. – К.: Академия, 2002. – 366 с.
- [Коджа, 2003] Т.И. Коджа. Автоматизированная система управления и контроля знаний в процессе обучения: автореф. дисс. канд. техн. наук: спец. 05.13.06 „АСУ и прогрессивные информационные технологии”, Одесса, 2003.
- [Снитюк, 2000] В.Е. Снитюк. Модели и методы определения компетентности экспертов на базе аксиомы несмещенности // Вестник Черкасского инженерно-технологического института, 2000, № 4, с. 121-126.
- [Снитюк, 2003] В.Е. Снитюк. Концептуальные принципы и методы проектирования систем автоматизированного контроля знаний // АСУ и приборы автоматики, 2003, вып. 123, с. 40-43.
- [Снитюк, 2008] В.Е. Снитюк. Оптимизация процесса оценивания в условиях неопределенности на основе структуризации предметной области и аксиомы несмещенности // Искусственный интеллект, 2008, № 3, с. 217-222.
- [Снитюк, 2010] В.Е. Снитюк. Элементы знаниеориентированных систем профессиональной подготовки адаптивного типа // Вестник Херсонского национального технического университета, 2010, № 2(38), с. 180-186.

---

## Информация об авторах



**Константин Юрченко** – Адъютант, Академия пожарной безопасности имени Героев Чернобыля, Ул. Оноприенко, 8, 18000, Черкассы, Украина; e-mail: yurchenko@gmail.com.  
Научные интересы: Компьютерные системы профессиональной подготовки.



**Виталий Снитюк** – Заведующий кафедрой, Черкасский государственный технологический университет, Бул. Шевченко, 460, 18006, Черкассы, Украина; e-mail: snytyuk@gmail.com.  
Научные интересы: Принятие решений в условиях неопределенности, эволюционное моделирование.

## ОРГАНИЗАЦИЯ ПРОЦЕССА ИНДИВИДУАЛИЗИРОВАННОГО ОБУЧЕНИЯ НА БАЗЕ АДАПТИВНОЙ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И КОНТРОЛЯ ЗНАНИЙ EDUPRO

Павло Федорук

**Abstract:** *The article is dedicated to solving a problem of differentiation and creation of individualized learning, that exist in modern distance education with the help of adaptive and intelligent technologies use. Developed technologies and methods let us realize practically the process of adaptive learning and knowledge control with intelligent Internet-technologies use.*

**Keywords:** *Adaptive Learning Systems, Intelligent Internet-technologies, Distance Learning.*

**ACM Classification Keywords:** *Education, Distance Learning, Adaptable Architectures.*

---

### Вступление

Методика преподавания с использованием технологий дистанционного обучения существенно отличается от традиционных технологий обучения и в основном опирается на самостоятельное изучение курса студентом, причем значительная часть работы преподавателя переводится на ЭВМ. Фактически обучающие системы "приобретают" знания эксперта-преподавателя и "доказывают" их к студенту. Поэтому можно констатировать, что основной особенностью дистанционного образования является предоставление студентам возможности самостоятельно получать необходимые знания, использовать современные информационные технологиями. Возможность индивидуализации обучения является одним из важнейших преимуществ использования информационных технологий в учебном процессе.

При существенной разнице в уровне базовой подготовки и индивидуальных способностей студентов одинаковый для всех план учебного процесса, принятый за основу в традиционных системах дистанционного обучения (СДО), является оптимальным в лучшем случае лишь для 15-30% студентов: для одних он слишком напряженный, для других, наоборот, недостаточно интенсивный. В результате неэффективно используются интеллектуальные и материальные ресурсы как индивида, так и общества.

Обзор наиболее известных систем широко используемых в современном дистанционном обучении классические подходы, таких как ANGEL, BlackBoard, Desire2Learn, ILIAS, Lotus LearningSpace, Moodle, WebCT, показал, что, как правило, учебный курс, представленный в среде СДО, представляет собой набор статических гипертекстовых документов. Все студенты получают одинаковый материал для изучения без учета их индивидуальных особенностей. Студент не может получить оперативную помощь от учителя или другого студента, как это происходит в обычном учебном классе. Новое исследовательское направление в сфере дистанционного обучения на Web-платформе - это адаптивные и интеллектуальные технологии в обучении [Brusilovsky, 1999], [Brusilovsky, 2003]. Задача этого направления исследований - включить в дистанционные обучающие системы возможности индивидуализации [Weber, 2001]. С помощью адаптивных и интеллектуальных технологий обучающая система учитывает индивидуальные способности студента, его предыдущие знания, умения. На основе этих данных о студенте процесс обучения проходит для него оптимальным путем [Gonzalez, 2003].

Все это способствует тому, что сегодня существует проблема обеспечения индивидуализированного обучения в дистанционном образовании, которая заключается в разработке методов, технологий и программных средств создания адаптивных систем дистанционного обучения на базе интеллектуальных Интернет-технологий [Федорук, 2006].

---

### Структура адаптивной обучающей системы

Для обеспечения организации процесса индивидуализированного обучения нами спроектирована структура адаптивной обучающей системы (рис. 1) [Федорук, 2009].

Студент взаимодействует с учебной системой через интерактивный модуль. Модуль обеспечивает возможность взаимодействия пользователей с системой с помощью интуитивно-понятного интерфейса. В процессе взаимодействия студента с системой студенческий модуль превращается в более совершенный, который точнее соответствует возможностям и потребностям студента и максимально точно отражает картину усвоенных знаний и приобретенных навыков. Использование модели студента позволяет спрогнозировать поведение студента и его мотивацию к обучению.

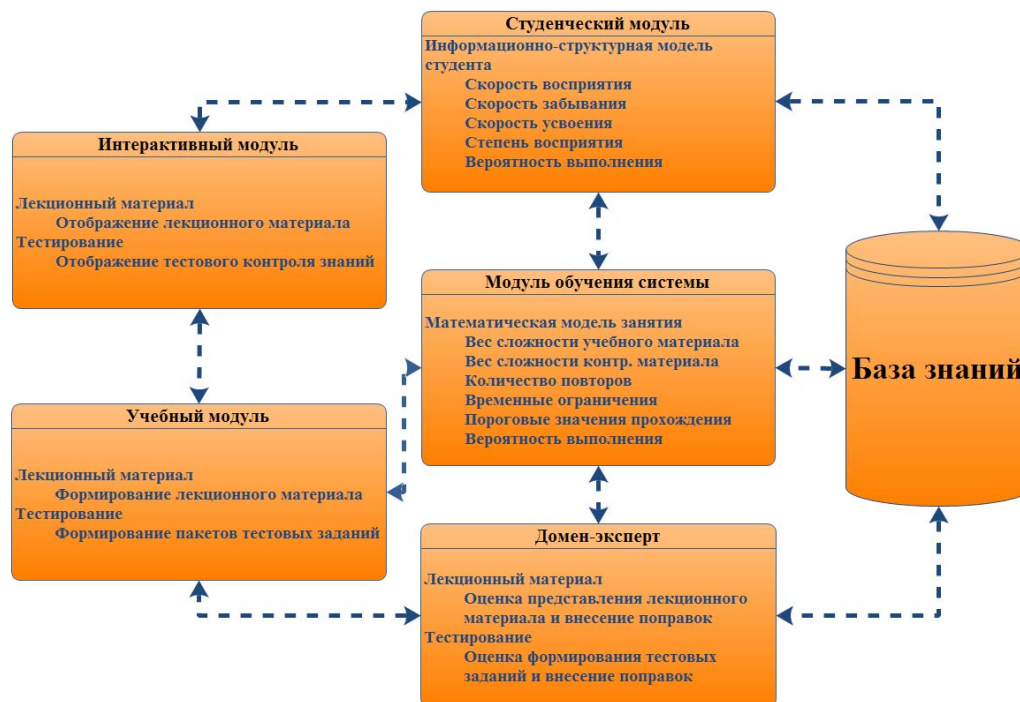


Рис. 1. Структура адаптивной обучающей системы

Домен-эксперт позволяет учебной системе функционировать в режиме эксперта (осуществлять адаптивный контроль за обучением). На него возлагается функция всесторонней оценки процесса обучения, качества знаний, прогресса и т.д.

Модуль обучения системы отвечает за процесс изменения учебного модуля. Этот процесс происходит под влиянием взаимодействия системы со студентом (т.е. индивидуально для каждого студента) по результатам, полученным от работы домена-эксперта.

Учебный модуль отвечает за создание и представление учебного контента средствами адаптивной навигации, адаптивного представления и адаптивного планирования курса.

Разработанные модели, технологии и методы построения адаптивных систем дистанционного обучения использованы нами при создании и внедрении в учебный процесс адаптивной системы дистанционного обучения и контроля знаний EduPro (рис.2).

### Построение адаптивной системы дистанционного обучения и контроля знаний EduPro

Для построения системы EduPro были применены интеллектуальные Интернет-технологии адаптивного представления, адаптивной навигации, адаптивного планирования курса, использованы средства относительно обеспечения доступа к базам данных, которые можно классифицировать по месту обработки данных внешними программами следующим образом [Федорук, 2009]:

- Средства, обеспечивающие доступ к базам данных на стороне сервера;
- Средства, работающие на стороне клиента.



Функционирование программной оболочки на стороне сервера обеспечивается СУБД PostgreSQL (v 8.x), сервером для обработки http запросов Apache (v 1.3.33 или выше), шлюзовым интерфейсом CGI (v 2.46 или выше), которые функционируют в операционной среде Unix.

На стороне сервера для доступа к базам данных используется механизм CGI (Fast CGI) и PHP.



Рис. 2. Схема построения системы EduPro

Все запросы для выбора нужной информации построены с использованием языка SQL. Интерфейсы пользователей разработаны на основе HTML-форм (рис. 3).

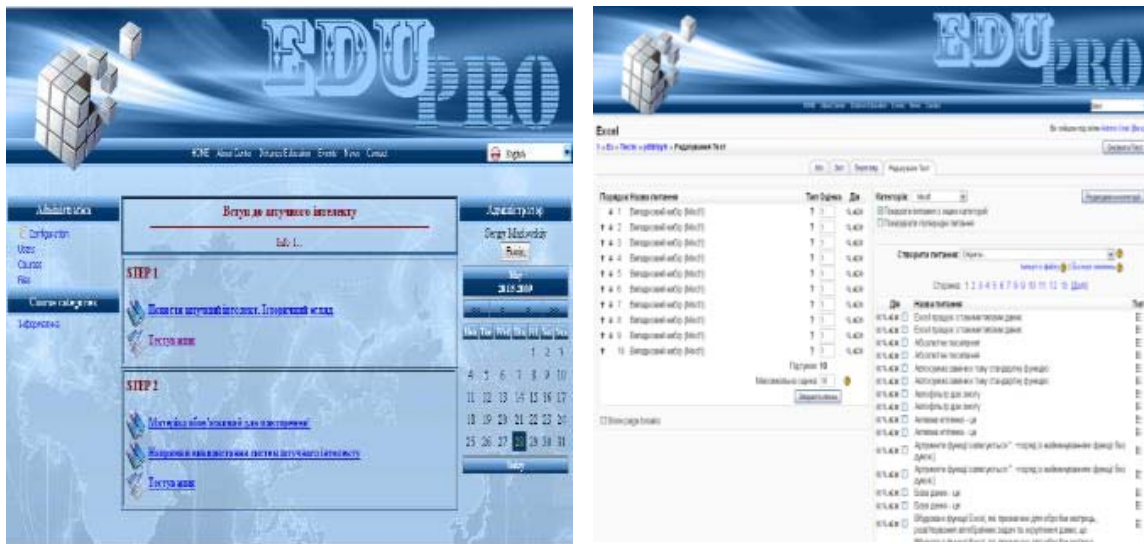


Рис. 3. Интерфейс и студента и преподавателя в системе EduPro

### Эффективность функционирования экспериментальной адаптивной системы дистанционного обучения и контроля знаний EduPro

Для сравнения классической и адаптивной систем дистанционного обучения проведено соответствующий эксперимент, в котором были задействованы 203 студента четвертого курса Прикарпатского национального университета им. Василя Стефаныка, которые проходили обучение по дисциплине "Искусственный интеллект" в двух учебных системах [Федорук, 2009].

Половина студентов училась в одной из самых популярных систем дистанционного обучения Moodle, другая - в адаптивной системе дистанционного обучения и контроля знаний EduPro.

Для обоснования выводов относительно отличия результатов обучения студентов в группах предварительно было подтверждено их первоначальную однородность успехов в учебе. Исследовались результаты обучения студентов за предыдущие годы. Проверку гипотез было проведено с использованием непараметрического критерия Манна-Уитни (реализованного в программе STATISTICA), поскольку исследуемые характеристики (результаты обучения) являются категориальным и измерены в порядковый шкале. Проверку однородности групп студентов, обучавшихся в системах EduPro и Moodle, проведено с помощью кластерного анализа (рис.4).

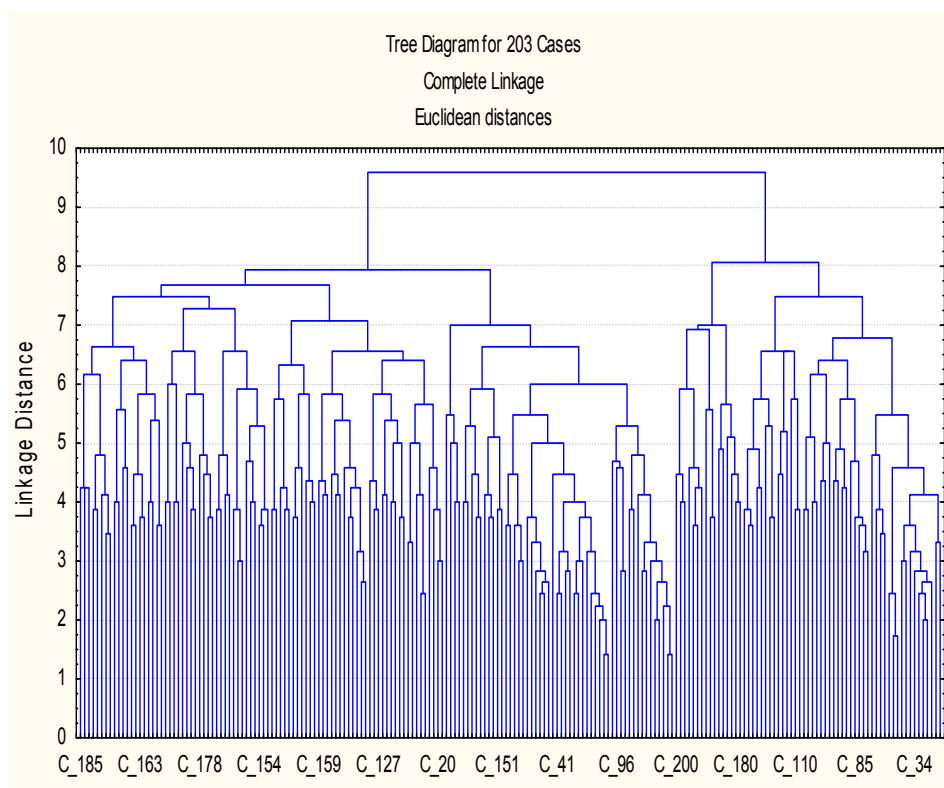


Рис. 4. Дендрограмма кластеризации студентов

Используя метод К-средних, было установлено, что наилучшим разбиением всех студентов на две группы, учитывая их различия в результатах обучения, является такое их разбиение, при котором первая группа содержит 122 студента из 203, вторая — 81. Из графика средних оценок (рис. 5) в полученных кластерах видно, что до первого кластера попали лучшие студенты, а второй формируют те студенты, которые учились хуже. К каждому кластеру попали как студенты из группы EduPro, так и из группы Moodle. Сравнение частот, с которыми представлены в данных группах лучшие студенты и студенты, которые учились хуже, позволило сделать вывод, что статистически достоверной разницы, которая влияла на результаты эксперимента между группами EduPro и Moodle по начальным уровнем способности студентов к обучению, нет ( $> 0, 05$ ).

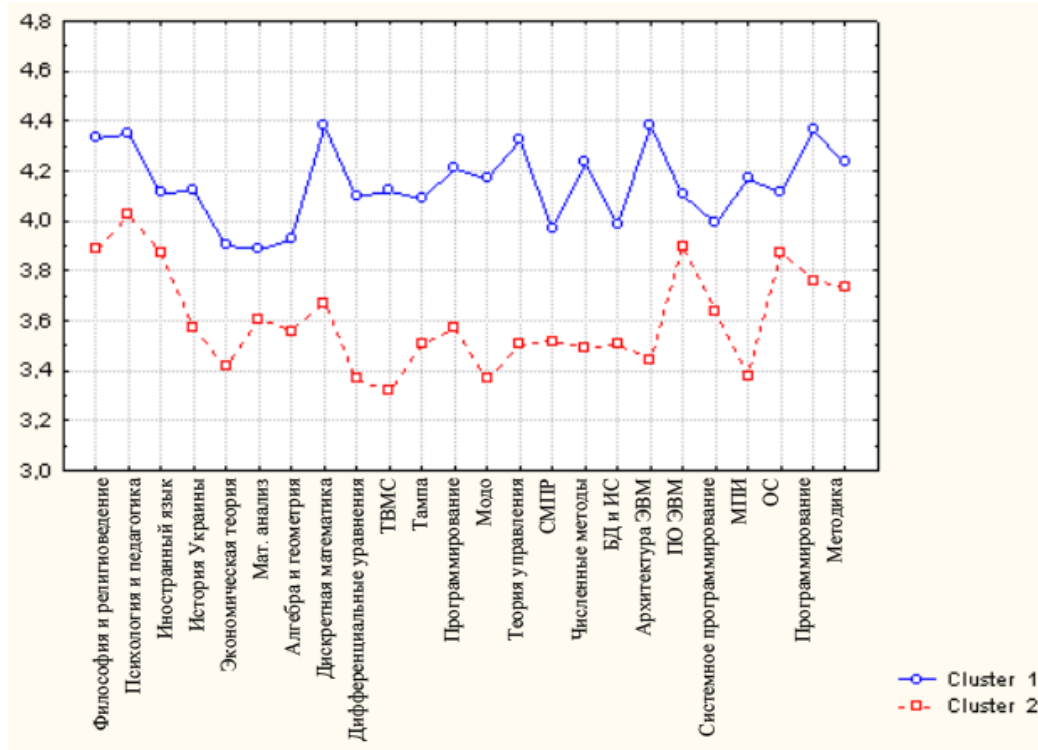


Рис. 5. График средних оценок

Во время проведения эксперимента определена эффективность функционирования системы EduPro по сравнению с системой Moodle для обучения студентов различных категорий (Cluster1, Cluster2). Результаты сравнения экспертных оценок в определенных группах студентов и итоговых экспертных оценок результатов обучения для всех студентов с использованием критерия Манна-Уитни позволили констатировать наличие существенной разницы между результатами всех показателей (рис.6).

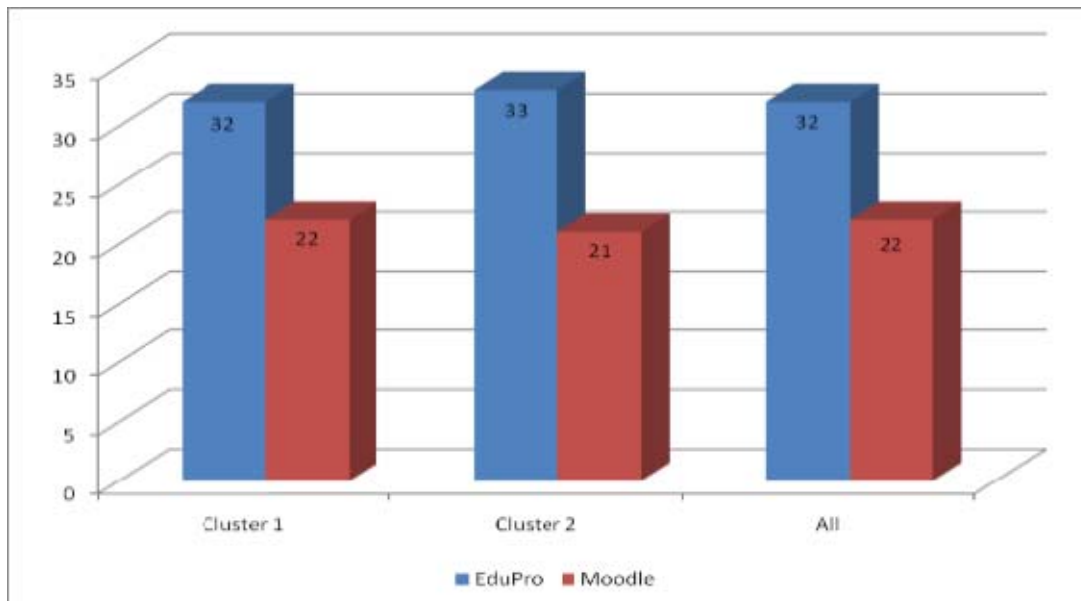


Рис. 6. Медианы экспертных оценок для студентов групп EduPro и Moodle

При фактически одинаковых количествах студентов в группах учащихся по разным системам, суммы рангов экспертных оценок у студентов группы EduPro значительно больше, чем у студентов группы Moodle. Особенно это касается результатов студентов, принадлежащих к другому кластеру. Это

свидетельствует, что использование механизмов индивидуализированного обучения в системе EduPro позволяет значительно улучшить качество усвоения знаний студентами, которые до этого обучались слабее.

Определив оценку математического ожидания и 95% доверительные интервалы для значений экспертной оценки в группах EduPro и Moodle, получено заключение, что средняя оценка в EduPro выше на 7,8 балла, чем в Moodle, что составляет 19,6% от максимально возможного балла (40).

Всего с надежностью 95% можно утверждать, что минимальное улучшение результатов в EduPro по сравнению с Moodle может составить 3,6 балла (8,9% от 40) и максимальное - 12,1 балла (30,3% от 40) (табл.1).

Таблица 1л Оценка математического ожидания и доверительные интервалы экспертной оценки знаний студентов по дисциплине "искусственный интеллект" (203 студенты)

Группы	Математическое ожидание	Доверительный интервал (95%)
Группа EduPro	30,0	(26,6; 33,5)
Группа Moodle	22,2	(21,3; 23,0)

### Вывод

На основе комплексного использования разработанных технологий и методов, применения интеллектуальных Интернет-технологий, построено адаптивную систему дистанционного обучения и контроля знаний EduPro и внедрено ее в учебный процесс. Доказана эффективность предложенных методов путем экспериментальных исследований эффективности функционирования разработанной системы. Показано, что применение систем дистанционного обучения позволяет не только сохранить качество традиционных технологий передачи знаний, но в ряде случаев за счет использования адаптивных алгоритмов добиться заметного повышения результатов обучения студентов. Сравнение результатов контроля знаний в группах студентов, которые учились с помощью одной из наиболее популярных систем дистанционного обучения, с группами, которые проходили обучение посредством адаптивной системы дистанционного обучения и контроля знаний EduPro, иллюстрирует более высокую оценку знаний у студентов в среднем на 19,6 %.

### Библиография

- [Brusilovsky, 1999] Brusilovsky P. Adaptive and Intelligent Technologies for Web-based Education. Kunstliche Intelligenz, 1999.
- [Brusilovsky, 2003] Brusilovsky P. Adaptive and Intelligent Web-based Educational Systems . International Journal of Artificial Intelligence in Education, № 13,2003.
- [Weber, 2001] Weber G. ELM-ART: An adaptive versatile system for Web-based instruction. International Journal of Artificial Intelligence in Education, № 12, 2001.
- [Gonzalez, 2003] Gonzalez C. M. A. Coaching Web-based collaborative learning based on problem solution differences and participation, International Journal of Artificial Intelligence in Education, № 13 (2-4), 2003.
- [Федорук, 2006] Федорук П. И. Использование адаптивных и интеллектуальных технологий в системах дистанционного обучения, УСиМ, № 5, 2006.
- [Федорук, 2009] Федорук П. И. Использование системы EduPRO для организации процесса адаптивного обучения. УСиМ, № 4, 2009.
- [Федорук, 2009] Федорук П. И. Исследование эффективности функционирования адаптивной системы дистанционного обучения EduPro. УСиМ, № 6, 2009.

### Authors' Information



**Pavlo Fedoruk** – *Precarpathian National Vasyl Stefanyk University, Director of Information Technologies Centre, Shevchenko str. 57, Ivano-Frankivsk 76025, Ukraine; e-mail: pavlo@pu.if.ua*

*Major Fields of Scientific Research: Mathematical and software support of calculated machines and systems, Artificial intelligence, Adaptive systems*

## ОБРАЗОВАНИЕ 2.0: ТЕНДЕНЦИИ И ТЕХНОЛОГИИ

Любовь Шарабаева, Ольга Захарова

**Аннотация:** В статье рассмотрены основные образовательные тенденции современного общества. Обоснована необходимость внедрения модели Образования 2.0 на основе технологий Web 2.0. Предложены основные подходы эффективной организации процесса обучения нового типа для повышения качества образования.

**Ключевые слова:** модели Образования, электронное обучение, технологии Web 2.0, системы совместного создания и редактирования контента

---

### Введение

Тенденции развития современного общества диктуют необходимость построения системы высшего образования инновационного типа. Это определяется как недостатками действующей системы образования – прежде всего, ее изоляцией от широкого социально-экономического контекста, так и решением вопроса о подготовке компетентных и креативных специалистов для реализации стратегии инновационного прорыва на мировые рынки.

Реформирование системы образования должно быть проведено в направлении укрепления ее связи с экономической и общественной жизнью общества, развитием и модернизацией экономики, соответствия современным потребностям рынка.

Инновационные образовательные проекты призваны реализовать современную образовательную парадигму, основным содержанием которой является глобализация, интернационализация и регионализация; технологизация и информатизация, перевод образования на новую методологическую и инструментальную базу; социализация; индивидуализация и личностно-ориентированное образование студентов, возрастание возможностей их творческой самореализации в образовательном процессе, синергия сотворчества преподавателей и студентов.

Современный вуз непременно должен стать обучающейся организацией, для которой процессы создания, накопления, использования и распространения знаний становятся ключевыми. Развитая способность к реализации таких процессов позволяет организации находиться в постоянной готовности совершенствовать свою деятельность.

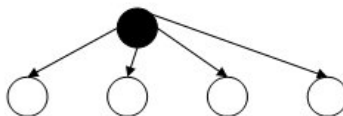
Характерными чертами обучающейся организации (Learning Organization), являются особенности организационной схемы, условий обмена информацией, стиля руководства, а также организационной культуры.

Реализация перечисленных аспектов наиболее полно осуществляется при внедрении системы управления знаниями (СУЗ) вуза, которая обеспечивает интегрированный подход к созданию, сбору, организации и использованию информационных ресурсов для решения задач обучения и доступа к ним с помощью интерактивной среды субъектов-участников образовательной деятельности [Шарабаева, 2008]. В связи с этим возрастает интерес к новым интеллектуальным технологиям, поддерживающим управленческие решения и способным с помощью аналитических и математических моделей адекватно реагировать на динамизм внешней среды. СУЗ вуза обеспечивает интеграцию разнородных источников знаний и их коллективное использование, а значит - позволяет решать наиболее актуальные задачи современной высшей школы: накопление и использование лучшего опыта в области организации административных процессов; моделирование и оптимизация процесса формирования основных компонент образовательного процесса в вузе; организация системы мониторинга качества образовательного процесса; управление процессом взаимодействия вуза с предприятиями-работодателями.

## Модели образовательных процессов

В научно-педагогической литературе на сегодняшний день можно встретить обобщенные модели образования вообще и электронного обучения, в частности, которые удобно распределить на три класса: модели Образования 1.0, 1.5 и 2.0 [Захарова, 2010].

Так, классическая модель Образования 1.0 представима в виде - один учитель и несколько учеников.



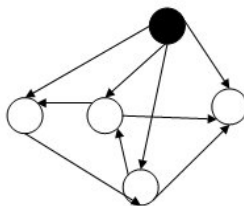
К недостаткам такой модели можно отнести следующие моменты:

- существование исключительно двух уровней - ученик и учитель, коммуникация между которыми осложнена;
- практически отсутствующее взаимодействие на одном уровне;
- недостаточная эффективность и функциональность форм организации учебного процесса.

В модели Образования можно выделить подмодель – Э-обучение, т.е. электронное обучение или обучение с использованием информационно-коммуникационных технологий. Основные способы организации образовательного процесса при Э-обучение 1.0 следующие:

- использование проектора, компьютера, интерактивной доски на семинарах/лекциях, что фактически совершенно не меняет модели обучения, а только добавляет наглядности к традиционным формам занятий;
- обучение через коммуникационную сеть, где отношения учитель-ученик сохраняются, поскольку преподаватель выкладывает свои материалы (тексты, лекции, презентации, задания, тесты) в компьютерную сеть, а студенты имеют возможность пользоваться ими при подготовке к семинарским занятиям и сдаче промежуточных и итоговых форм контроля. Компьютер и Интернет в этом случае играют лишь роль транслятора, передатчика информации из одного источника к другому и совершенно не использует даже основной части своих возможностей.

К формату переходной модели Образования 1.5 - один учитель и несколько учеников – добавляется также и горизонтальная коммуникация.



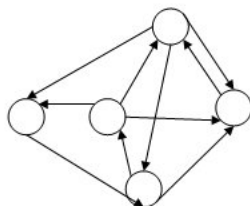
Особенностями такой модели являются:

- обучение в группе;
- взаимопомощь;
- возможность для каждого быть и учеником, и учителем для других учеников;
- обеспечение свободного передвижения информации и материалов;
- существование уровня выше – учитель, читающий лекции и принимающий экзамены.

Подмодель Э-обучение 1.5 характеризуется подключением к существующей структуре также и обратной связи. В числе особенностей Э-обучения 1.5 можно выделить следующие:

- появление таких возможностей как форумы, групповые обсуждения в процессе обучения;
- обеспечение обратной связи с преподавателем;
- увеличение скорости распространения информации;
- сокращение времени реагирования в сети.

Модель Образования 2.0 подразумевает обучение в группе, где каждый является специалистом в каком-либо вопросе и хочет узнать нечто другое.



Образование 2.0 – это совокупность «электронных» способов доступа, анализа, обработки и отклика на индивидуально/коллективно обрабатываемые данные, представляющие образовательный интерес для различных пользовательских групп. Современные технологии вносят интерактивность в процесс обучения.

В модели Образования 2.0 отсутствует четкое разделение «преподаватель-студент», условно можно выделить роли: «наставник» и «эксперт». Модель позволяет отойти от жестко фиксированного плана обучения. Вместо него используется понятие индивидуальной образовательной траектории, маршрутизации обучения, когда студент сам выбирает, какие темы ему нужно изучить для того, чтобы достичь необходимой компетенции в интересующей его области знаний. Такая модель нового типа предусматривает наличие информационно-образовательной среды, содержащей значительное количество различных инструментов. Это могут быть учебные пособия в их классическом понимании, виртуальные кафедры и лаборатории, видеоролики, компьютерные деловые игры и т.д. В перспективе разрабатывается «идеальный» персональный курс для каждого обучающегося. Основная задача Образования 2.0 - создать инфраструктуру, которая придаст образовательным материалам живость, откроет возможность перманентного обновления и обеспечит к ним доступ в режиме 24x7.

Главный акцент Образования 2.0 – передача инициативы в управлении знаниями самим студентам, самоорганизация обучения, – будет понят только тогда, когда студенты будут готовы эту инициативу проявить. В этом случае очень подходит известная фраза: готов ученик, готов и учитель.

Среди основных особенностей Э-обучения 2.0 можно отметить:

- возможность обучения в группе;
- взаимопомощь;
- активная коммуникация и широкое распространение информации;
- формирование сообществ по «интересам»/целям, т.е. тематических сообществ;
- наличие возможности индивидуального обучения в соответствии с выбранным маршрутом и уровнем сложности изучаемого курса;
- свобода выбора;
- самопроизводство и самопотребление информации.

Модель Образование 2.0 неразрывно связана с Э-обучением 2.0, реализуемым на базе Web 2.0.

---

### Электронное обучение

Необходимость формирования систем управления знаниями возникла, прежде всего, в вузах, активно внедряющих в образовательный процесс технологии электронного обучения (e-Learning). Поскольку электронное обучение базируется на использовании электронных образовательных ресурсов, доступных в сетях Интранет и Интернет, это ведет к созданию электронного контента, формированию электронных библиотек и распределенных образовательных сетей, объединенных единой информационной системой; использованию системы управления учебным процессом (LMS) и системы управления контентом (LCMS). СУЗ вуза позволяет объединить эти ресурсы в единую управляющую систему.

Сегодня необходимость применения технологий электронного обучения уже не является предметом дискуссии - это вызов времени, принятый всеми субъектами образовательной сферы. В перспективе – развитие мобильного обучения (m-Learning) и, так называемого, вездесущего, всепроникающего электронного обучения u-Learning (ubiquities learning). Внедрение всепроникающего обучения позволит учащимся с помощью беспроводной связи и специальных технологий в любом месте загружать и изучать перманентно обновляемый образовательный контент высокого качества.

Совершенствование технологий электронного обучения служит прочной основой все более широкого распространения дистанционной формы обучения для различных уровней образования. Под дистанционным обучением (ДО) понимается такое взаимодействие преподавателя и учащихся, а также учащихся между собой на расстоянии, которое отражает все присущие учебному процессу компоненты (цели, содержание, методы, организационные формы, средства обучения), с помощью специальных средств Интернет-технологий и интерактивных технологий.

На наш взгляд, технологии электронного обучения должны быть использованы в образовательном процессе современного вуза для всех форм и уровней обучения, и, прежде всего, с целью изменения образовательного процесса в контексте основных актуальных образовательных концепций: непрерывность во времени и пространстве; дистанционность и мобильность; проектность и технологичность; интерактивность и сотрудничество - учебный процесс как совместное творчество и диалог преподавателя и студента; ориентация учебного процесса не только на содержание и логику предмета, но и на логику восприятия в условиях информационно-коммуникационного взаимодействия; соблюдение основного принципа вебдвухольных технологий – принципа наследования.

Для эффективной организации процесса обучения нового типа и повышения качества образования необходимо гармоническое сочетание таких трех аспектов, как:

- технический, предоставляющий удобную среду для организации учебного процесса и эффективного интерактивного взаимодействия;
- организационный, определяющий типы взаимосвязи и права участников образовательного процесса;
- информационный, отвечающий за наполнение системы контентом и перманентное, динамичное его обновление.

Современный образовательный процесс должен рассматриваться как проект создания нового уникального контента и связывания различных его составляющих между собой. Реализация такого проекта требует наличия удобной среды, позволяющей участникам проекта совершать любые действия относительно представления, обсуждения и классификации публикуемого материала. Для того чтобы обеспечить такие возможности, необходимо прибегнуть к бурно развивающимся на сегодняшний день и активно внедряющимся во все сферы жизнедеятельности человека, технологиям Web 2.0.

---

### **Технологии Web 2.0 в образовании**

---

Новый термин – Web 2.0 был введен в статье Тима О'Рейли 30 сентября 2005, в которой он рассказал о новой концепции Web, дав следующее определение «Web 2.0: Web - как платформа», рассматривая Web 2.0, в первую очередь, как множество правил и практических решений для ведения эффективного, социально- ориентированного бизнеса.

Понятие Web 2.0 многогранное и трудно формулируемое, в связи с чем, его стоит рассматривать не только с позиции бизнеса, но и с позиции образовательного процесса и образовательной системы в целом.

Эволюция Web привела к смене принципов взаимодействия пользователей с ресурсами. В основе Web 2.0 лежат не ресурсы, а пользователи, их знания, их взаимодействие. Веб-документ (веб-страница) становится веб-приложением, а Интернет и все предоставляемые им возможности, начиная с поиска (поисковой машины) – услугой.



Под веб-приложениями понимаются приложения, функциональные возможности которых обеспечиваются сервером и доставляются конечным пользователям по сети, такой как Интернет или Интранет.

Среди основных веб-двухпольных технологий следует выделить технологию Ajax (*Asynchronous JavaScript*), – технологию создания динамических интерфейсов веб-приложений, основанную на асинхронном фоновом обмене данными в XML формате между сценариями написанными на JavaScript и сервером. Применение Ajax технологии позволяет достичь повышения удобства работы, простоты использования (*usability*) и расширенной функциональности разрабатываемых веб-приложений, стирая различия между обычными и веб-приложениями [Форсье Дж. и др., 2010].

Таким образом, если с Web 1.0 в сети мы имели возможность пользоваться только одиноко отстоящими и никак не связанными между собой сайтами со статичной информацией, и такого же типа форумами для общения и обсуждения важных проблем, то сегодня им на смену вместе с концепцией Web 2.0 пришли совершенно новые способы коммуникации, в частности:

- Wiki-системы - гипертекстовые среды (веб-сайт) для сбора и структурирования письменных сведений, которые характеризуются такими признаками, как: множество авторов; возможностью многократно править текст посредством самой среды (веб-сайта), без применения приложений на стороне редактора; возможностью наблюдать за динамикой изменений сразу после их внесения; разделять информацию на порции, имеющие свое название (катеорию); особым языком разметки, позволяющим легко и быстро размечать в тексте структурные элементы, форматирование, гиперссылки и т. п.; учитывать изменения (контроль версий) текста и возможность отката к ранней версии документа.

- Системы совместного создания и редактирования контента (*collaboration editing*), например, GoogleDocs, GoogleWave и др., обеспечивающие возможность совместного одновременного доступа к документу, работу с ним и его обсуждение.

- Живой блог (англ. *blog*, от *Web log*, виртуальный «сетевой журнал или дневник событий») – веб-сайт, основное содержимое которого – регулярно добавляемые записи, изображения или мультимедиа. По авторскому составу блоги могут быть личными, групповыми (корпоративные, клубные) или общественными (открытыми). По содержанию – тематическими или общими. Так же блоги могут объединяться в сети блогов, по тематическим признакам или по другим критериям. Для блогов характерна возможность ознакомления с материалом одновременно многими пользователями сети Интернет и оставления отзывов к записям непосредственно в электронной среде; публичность. Эта возможность позволяет использовать блоги в качестве среды сетевого общения, имеющей ряд преимуществ перед электронной почтой, новостными группами, веб-форумами и чатами. Появление и быстрое распространение блогов вписывается в концепцию Web 2.0, создавая так называемую «редактируемую Паутину» (*writable Web*). В новых интернет-проектах существует также аббревиатура UGC – *User Generated Content* (контент, создаваемый самими пользователями). Живость блогу придает именно способность обновлять свое содержимое без вмешательства пользователей (технология Ajax).

- RSS - семейство XML-форматов, предназначенных для описания лент новостей, анонсов статей, изменений в блогах и т.п. Информация из различных источников, представленная в формате RSS, может быть собрана, обработана и представлена пользователю в удобном для него виде специальными программами-агрегаторами либо самим браузером. Первоначально эта технология использовалась на новостных ресурсах и в блогах, но постепенно сфера применения расширилась, и теперь подписка на свежий информационный контент может быть искусно внедрена и в образовательный процесс.

- Вебинар – он-лайн семинар, лекция, курс, презентация, организованный при помощи web-технологий, целью которого является максимально эффективно обеспечить доставку контента к слушателю. Проведение вебинара позволяет охватить широкую аудиторию не только пассивных участников, но и дает возможность проявить желающим активность. Вебинар как современный телекоммуникационный инструмент образовательного процесса достаточно перспективен, т.к. прежде всего вызывает огромную заинтересованность у участников в его проведении.

Все названные технологии и открывающиеся, благодаря их использованию, возможности, способствуют формированию сетевых тематических сообществ и их росту, подобно большим городам, которые самоорганизуются на удобных площадках. Активность современных сетевых сообществ настолько велика, что журнал «Тайм» в 2006 году назвал сетевого автора человеком года и поместил на обложку своего журнала слово «You» (Вы). Основной особенностью является то, что почетный титул заслужил не потребитель – читатель и зритель каналов, а тот, кто сам активно участвует в формировании содержания. В формат нового типа непременно должны перестроиться и все образовательные отношения в вузе для повышения качества обучения и достижения синергетических эффектов благодаря коллективному взаимодействию.

---

### Заключение

На основе проведенного анализа современных образовательных тенденций и ключевых технологий Web 2.0 можно заключить, что представляется не только целесообразным, а и неизбежно необходимым внедрять в образовательный процесс интеллектуальные Web-технологии, подключать ресурсы и использовать возможности, предоставляемые Интернет-средой. Это позволяет активно вовлекать в образовательный процесс всех его участников, повышать мотивацию обучения, и соответственно качество образования.

Реализация модели Образования 2.0 на практике требует использования и творческого осмысления бурно развивающихся на сегодняшний день и активно внедряющихся во все сферы жизнедеятельности человека технологий Web 2.0.

---

### Библиография

- [Захарова, 2010] О.В. Захарова. Модели перехода экономико-образовательных процессов к веб-двухмерной концепции поддержки знаний. Тезисы докладов II международной научно-практической конференции Современные проблемы моделирования социально-экономических систем. ИД ИНЖЭК, Харьков, 2010.
- [Форсье Дж. и др., 2010] Дж. Форсье, П. Биссекс, У. Чан. Django. Разработка веб-приложений на Python. Символ-Плюс, СПб, 2010.
- [Шарабаева, 2008] Л.Ю. Шарабаева. Теоретические основы и ключевые технологии построения системы управления знаниями вуза. Управленческое консультирование. №2, 2008.

---

### Информация об авторах



**Любовь Шарабаева** – к.ф.-м.н., доцент кафедры информатики Северо-Западной академии государственной службы, адрес для переписки: пр. Авиаконструкторов, д.116, кв.36, Санкт-Петербург-197373, Россия; e-mail: [shar\\_lu@mail.ru](mailto:shar_lu@mail.ru)

Основные направления научных исследований: E-learning, системы управления знаниями, онтологический инжиниринг, математическое моделирование управления обучением и контроля знаний, интегрированные экспертные системы



**Ольга Захарова** – преподаватель кафедры экономической кибернетики Харьковского национального экономического университета, адрес для переписки: ул. Мироносицкая, 93-А, кв. 4, Харьков-61023, Украина; e-mail: [harizmalife@gmail.com](mailto:harizmalife@gmail.com)

Основные направления научных исследований: E-learning, Datamining, искусственный интеллект, информационные системы и технологии, нелинейная динамика, моделирование и прогнозирование макроэкономических процессов в условиях нестационарной среды

## МУЛЬТИМЕДИЙНЫЕ ОБРАЗОВАТЕЛЬНЫЕ СИСТЕМЫ: ОПЫТ ВНЕДРЕНИЯ

Тамара Клебанова, Роман Яценко, Ольга Захарова

**Аннотация:** В статье рассмотрен опыт внедрения мультимедийной образовательной системы, включающей образовательный портал, систему электронной кафедры и адаптивную систему дистанционного обучения. Приведена структура и функциональные задачи каждой составляющей. Проанализированы статистические показатели, характеризующие результаты внедрения мультимедийных технологий в образовательную систему.

**Ключевые слова:** образовательная система, мультимедиа, информационные технологии, дистанционное обучение, вебинар.

### Введение

Современный этап развития мировой экономики характеризуется переходом от индустриального к информационному обществу во всех сферах человеческой деятельности. Такой переход требует внедрения интеллектуальных информационных технологий, построенных на новых принципах массовой переработки и использования информации. Основная задача подобных технологий – создавать информационную среду для аккумуляции и продуцирования новых знаний. В связи с этим большое значение приобретают вопросы повышения качества, эффективности и содержания образования в высшей школе. Развитие мультимедийных информационных технологий влечет за собой необходимость серьезной перестройки образовательных систем управления высшей школы и, прежде всего, системы управления кафедрой.

### Структура и функциональные задачи составляющих системы

В качестве структурных составляющих предлагаемой системы рассматриваются: образовательный портал кафедры, система электронной кафедры и адаптивная система дистанционного обучения (рис. 1).

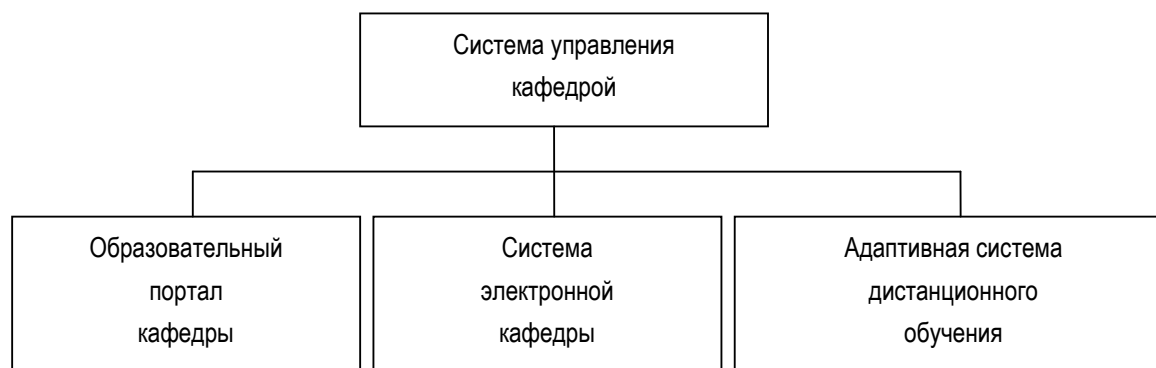


Рис. 1. Схема системы управления кафедрой

Ниже приведен состав функциональных задач каждой составляющей системы. Так, назначением образовательного портала кафедры является [Клебанова, 2009]:

- создание электронных учебно-методических материалов по различным дисциплинам, в том числе тестов для разных форм контроля;
- доступ к виртуальным образовательным учреждениям и доставка образовательной информации;
- организация самоконтроля и тестирования, аттестации;

- создание персональной образовательной среды;
- текущий контроль успеваемости студентов при работе с учебно-методическими материалами;
- организация удаленного консультирования;
- управление образовательным процессом в дистанционной форме;
- реализация систем адаптивного дистанционного обучения;
- применение эффективных схем управления образовательными учреждениями;
- проведение вебинаров на основе использования современных мультимедийных телекоммуникационных инструментов;
- использование систем коллективной работы, вики-документов для создания и хранения образовательного контента.

Информационная модель образовательного портала приведена на рис. 2.



Рис. 2. Информационная модель образовательного портала

К основным функциональным задачам электронной кафедры относятся: формирование базы знаний кафедры, электронный документооборот кафедры, управление научной деятельностью кафедры, управление учебным процессом, контроль успеваемости студентов, маркетинговая деятельность. В базе знаний кафедры хранятся:

- Перечень дисциплин по специальностям;
- Структурно-логические схемы специальностей;
- Библиотека программ курсов, учебников, учебных пособий и конспектов лекций по дисциплинам специальностей;
- Дополнительная литература и ссылки на внешние ресурсы.

Электронный документооборот позволяет значительно снизить трудоемкость выполнения отдельных работ на кафедре и включает: протоколы заседаний; учебные планы; нагрузку по кафедре; служебные записки; справочную информацию; отчеты кафедры; базу данных методических указаний по дисциплинам кафедры; презентации; брошюры для проведения профориентационной работы кафедры; другое.

Научная деятельность кафедры находит отражение в следующих разделах: публикации преподавателей; тематика госбюджетных и хоздоговорных работ кафедры; информация по аспирантам кафедры; отчеты по науке; конференции; гранты кафедры.

При управлении учебным процессом рассматриваются следующие задачи: формирование расписания занятий и консультаций преподавателей кафедры; использование средств коммуникации (дистанционный офис); разработка системы контроля качества исполнения заданий по кафедре; определение внутреннего рейтинга преподавателей.

Контроль успеваемости студентов осуществляется по направлениям: формирование электронного журнала посещаемости занятий студентами всех курсов специальностей кафедры; оценка текущей успеваемости по дисциплинам учебного плана специальностей; анализ результатов блочно-модульного контроля; анализ результатов экзаменационных сессий.

Маркетинговая деятельность кафедры включает: анализ работы портала; реклама специальностей; профессионально-ориентационная работа (информация для абитуриентов на портале кафедры; заключение договоров с учебными заведениями I-II уровня аккредитации и анализ их выполнения; проведение презентационных мероприятий кафедры).

К числу наиболее важных составляющих системы управления кафедрой относится адаптивная система дистанционного обучения (АСДО), элементный состав которой приведен в табл. 1.

Таблица 1

Элементный состав и перечень функциональных задач АСДО

Элементы	Задачи
1. Управление учебным процессом	Описание курса Календарь Объявления Учебный план
2. Библиотека курса	Лекции Литература Глоссарий Методические указания
3. Контроль знаний	Тесты Задания
4. Совместная работа	Wiki Работа в группах
5. Коммуникация	Форум Служба сообщений Чат
6. Мониторинг	Доступ к курсу Учебные материалы Средства коммуникации Контроль знаний

Назначением первого элемента является управление учебным процессом на основе использования календарного планирования и учебных планов. «Библиотека курса» предназначена для доставки студентам различных учебных материалов (лекции в любой форме представления, методические указания, глоссарий, дополнительная литература и т.д.). С помощью элемента «Контроль знаний» осуществляется оценивание результатов обучения на базе тестов и индивидуальных заданий. Совместная работа между студентами реализуется в виде работы в группах и создания Wiki, представляющего собой инструмент аккумуляции знаний в энциклопедической форме. Коммуникация как между преподавателем и студентами, так и между студентами основана на использовании таких

общепринятых на сегодня инструментов, как форум, чат и рассылка сообщений. Последним элементом АСДО является блок «Мониторинг». Схема взаимосвязи элементов АСДО приведена на рис. 3.

Адаптивная система дистанционного обучения имеет ряд преимуществ [Федорук, 2008]:

- уменьшение непроизводительных затрат труда преподавателя;
- свободный выбор студентом стратегии и тактики обучения;
- непрерывная обратная связь в процессе обучения;
- оперативность и объективность контроля и оценки результатов обучения;
- индивидуализация учебной деятельности;
- дифференцированный подход к студентам;
- повышение мотивации обучения.

К числу важнейших элементов АСДО относится система мониторинга, назначением которой является сбор, обработка и анализ информации о состоянии учебного процесса. В табл. 2 по каждому элементу учебного процесса приведен перечень выходных документов и показателей, на основе которых проводится анализ их состояния.

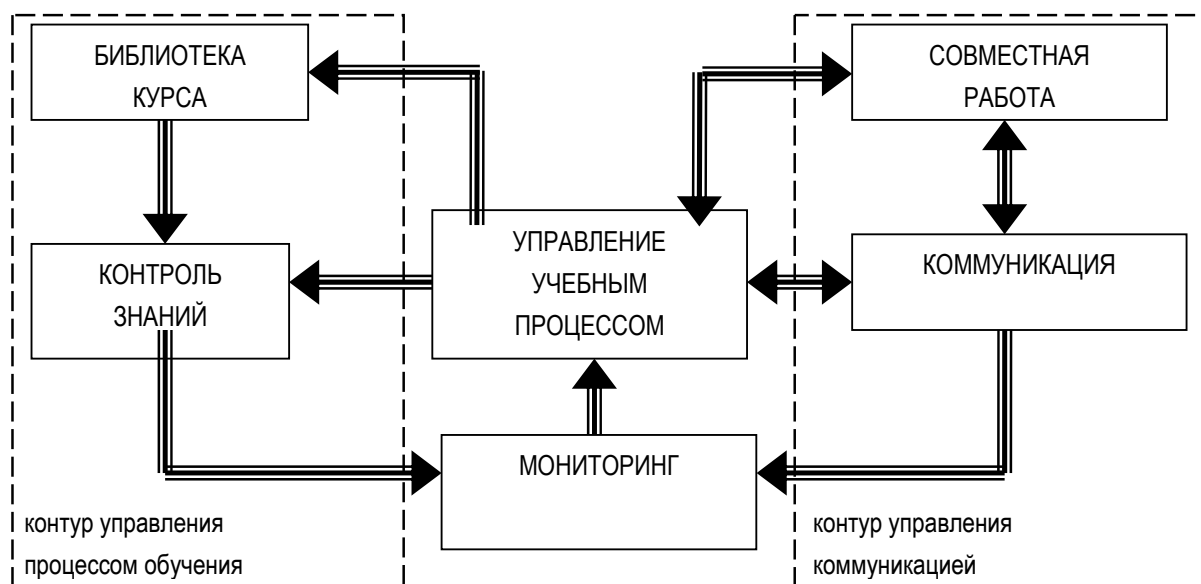


Рис. 3. Схема взаимосвязи контуров управления и элементов АСДО

Таблица 2

Список выходных документов ( журналов ) по элементам учебного процесса

Элемент	Журналы	Показатели
Доступ к курсу	Детали трафика	Количество обращений за час, день, месяц, год
Учебные материалы	Статистика по библиотеке Статистика персонального доступа	Количество обращений к материалам библиотеки Количество обращений Время последнего обращения

Средства коммуникации по курсу	Активные темы форума	Количество ответов в теме
	Часто просма-триваемые темы	Количество просмотров темы
	Последние активные темы	Время последнего сообщения в теме
Контроль знаний студента по темам курса	Статистика по пользователю в разрезе курса	Худшая, лучшая и средняя оценки
		Количество попыток
		Среднее время
	Статистика по вопросам темы	Время последней попытки
		Худшая, лучшая и средняя оценки
	Статистика по вопросу темы	Ожидаемый вариант ответа
		Распределение вариантов ответов на вопрос

Анализ приведенных показателей позволил осуществить разработку рекомендаций по дальнейшему совершенствованию образовательной системы.

#### Опыт внедрения адаптивной системы дистанционного обучения

Значительный интерес представляет анализ некоторых показателей адаптивной системы дистанционного обучения, которые были получены в результате ее внедрения в 2009-2010 учебном году кафедрой экономической кибернетики Харьковского национального экономического университета.

На рис. 4 приведена динамика изменения количества посетителей образовательного портала за период с сентября по декабрь месяцы 2009 года. Как видно, наблюдается рост количества посещений образовательного портала, что свидетельствует об устойчивом интересе студентов к различным технологиям е-образования. Так, в декабре этот показатель вырос по сравнению с сентябрем более чем в 4 раза.

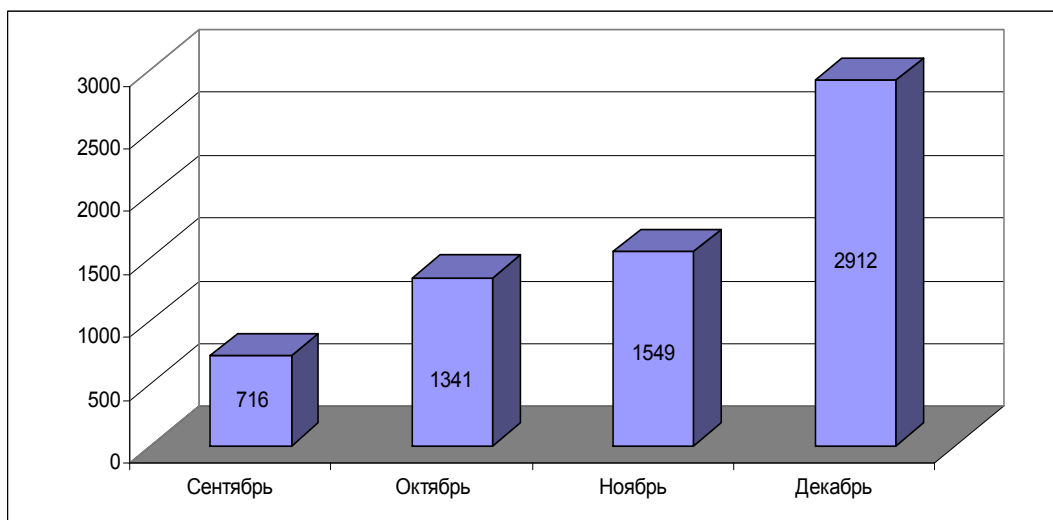


Рис. 4. Количество посетителей образовательного портала за осенний семестр 2009 г.

На рис. 5 приведена диаграмма, которая позволяет определить каким элементам АСДО студенты отдают предпочтение в процессе обучения по анализируемым курсам.

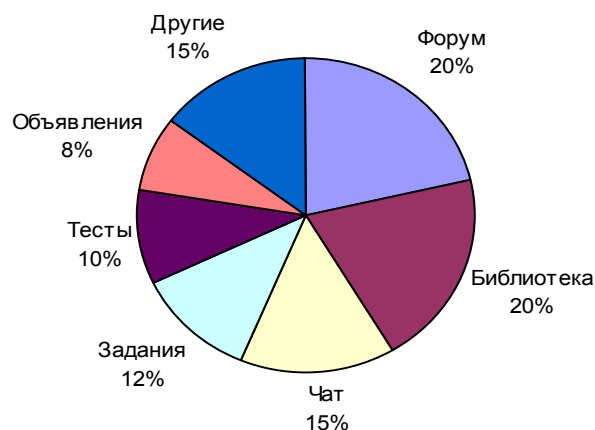


Рис. 5. Диаграмма предпочтения элементов АСДО

Диаграмма показывает, что основное внимание следует уделять дальнейшему развитию библиотек курсов (20%) и средствам коммуникации (35%). Необходимо отметить, что применение различных технологий дистанционного обучения приводит к выравниванию учебной нагрузки студентов в течение семестра (рис. 6), что значительно влияет на качество обучения.

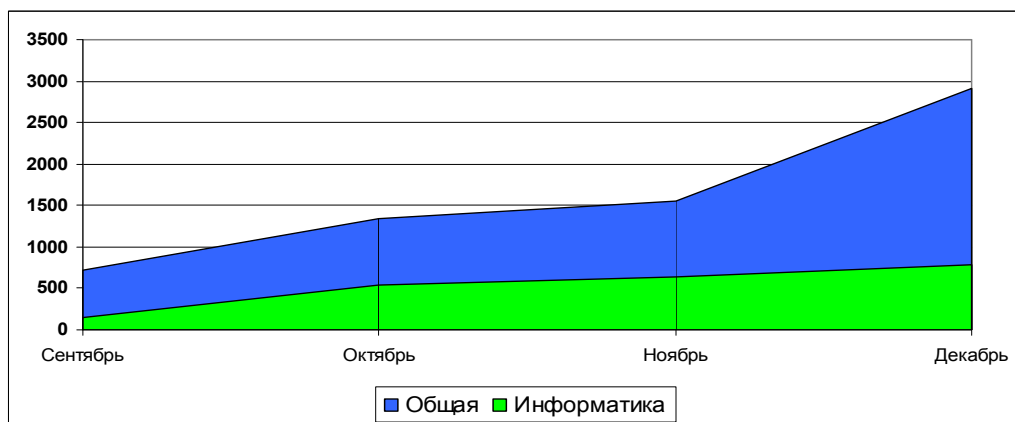


Рис. 6. Количество обращений студентов к курсам

Последнее объясняется тем, что учебная дисциплина в дистанционном обучении - понятие гораздо более конкретное, чем при очных занятиях. Студент должен ответить на тесты, знать сроки исполнения заданий, учесть требования по посещению конкретного дистанционного курса и т.д. Имеется жесткий график учебной отчетности, контроля, общения с преподавателем и администрацией программы. Любые отклонения от этого графика вызывают мгновенную реакцию преподавателя, который, либо применяет санкции, либо предлагает помощь студенту. Анализ персональной статистики (табл. 3) дает возможность



определить основные направления дифференцированного подхода к студенту, сформировать гибкие индивидуальные учебные планы.

Большое значение в дальнейшем развитии АСДО приобретают современные телекоммуникационные инструменты – вебинары [Захарова, 2010]. Под последними понимается он-лайн семинар, лекция, курс, презентация, организованный при помощи мультимедиа-технологий, что позволяет охватить широкую аудиторию слушателей, повысить интерактивность и скорость реагирования участников мероприятия.

Цель вебинара – обеспечить доставку контента к студенту, показать пути эффективного использования полученных знаний. На рис. 7 приведена динамика посещения вебинаров за декабрь 2009 г., которая свидетельствует о повышенном интересе как студентов, так и преподавателей к этому телекоммуникационному инструменту.

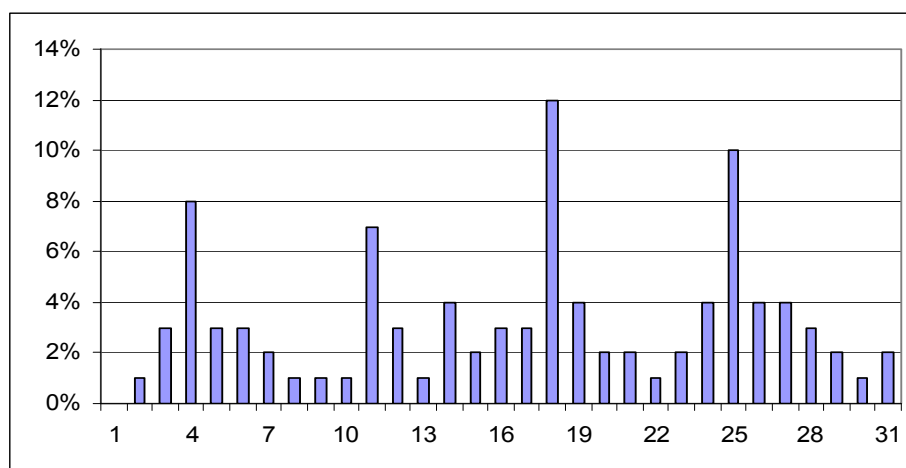


Рис .7. Доля обращений к вебинарам за день от общего числа за месяц (%)

Так, общее число обращений к вебинарам за декабрь месяц составило 1000 раз, зарегистрировалось более 60 человек (преподавателей и студентов), активно участвовали в дискуссиях он-лайн около 30 человек. Протоколы вебинаров показали, что они обладают рядом преимуществ, основными из которых являются:

- эффективная интерактивная форма представления знаний;
- удобная диалоговая форма взаимодействия между участниками;
- существенное увеличение скорости обмена мнениями и знаниями во время диалога;
- продуцирование новых знаний в форме мультимедиа-контента всеми участниками в процессе коллективной работы.

## Заключение

Внедрение на основе мультимедийных информационных технологий всех элементов системы управления кафедрой в полном объеме и, прежде всего, адаптивной системы дистанционного обучения значительно повысит эффективность её работы, заинтересованность студентов в процессе обучения, позволит усовершенствовать подготовку компетентного специалиста на основе использования технологий е-образования.

---

## Библиография

---

- [Захарова, 2010] О.В. Захарова. Модели перехода экономико-образовательных процессов к веб-двухканальной концепции поддержки знаний. Тезисы докладов II международной научно-практической конференции Современные проблемы моделирования социально-экономических систем. ИД ИНЖЭК, Харьков, 2010.
- [Клебанова Т.С. и др., 2009] Т.С. Клебанова, Р.Н. Яценко. Современные информационные технологии поддержки дистанционного обучения. Тезисы докладов I международной научно-практической конференции Современные проблемы моделирования социально-экономических систем. ИД ИНЖЭК, Харьков, 2009.
- [Федорук, 2008] П.И. Федорук. Адаптивная система дистанционного обучения и контроля знаний на базе интеллектуальных Интернет-технологий. ИД Прикарпатского национального университета, Ивано-Франковск, 2008.

---

## Информация об авторах

---



**Тамара Клебанова** – д.э.н., зав. кафедры экономической кибернетики Харьковского национального экономического университета, e-mail: [t\\_kleb@ukr.net](mailto:t_kleb@ukr.net)

*Основные направления научных исследований: методы анализа и своевременного предупреждения дестабилизаций функционирования социально-экономических систем, методы экономико-математического моделирования экономических проблем*



**Роман Яценко** – к.э.н., доцент кафедры экономической кибернетики Харьковского национального экономического университета, e-mail: [roma\\_yac@mail.ru](mailto:roma_yac@mail.ru)

*Основные направления научных исследований: моделирование ценовой политики предприятия, методы компьютерного образования*



**Ольга Захарова** – преподаватель кафедры экономической кибернетики Харьковского национального экономического университета, адрес для переписки: ул. Мироносицкая, 93-А, кв. 4, Харьков-61023, Украина; e-mail: [harizmalife@gmail.com](mailto:harizmalife@gmail.com)

*Основные направления научных исследований: E-learning, Datamining, искусственный интеллект, информационные системы и технологии, нелинейная динамика, моделирование и прогнозирование макроэкономических процессов в условиях нестационарной среды*

## «ОТКРЫТЫЕ» ОТВЕТЫ В КОМПЬЮТЕРНОМ ТЕСТИРОВАНИИ ЗНАНИЙ

Леонид Леоненко

**Аннотация:** Обосновывается возможность адекватной оценки компьютером ответов “свободной” формы, в том числе на натуральном языке. Методы сравнения ответов, позволяющие оценивать “открытые” тестовые задания, основаны на выводах по аналогии. Для разных типов заданий применяются разные математические меры оценки «аналогичности» ответов на них. Наиболее важными являются меры, основанные на теории подобия символьных последовательностей, позволяющие адекватно оценивать лексическое и структурное сходство текстов естественных языков. Система компьютерного тестирования знаний «Control» использует «библиотеку» мер аналогичности различных видов, предоставляющую преподавателю средства оценки ответов, предъявленных в разнообразных семантических формах (причем эта библиотека может пополняться новыми типами указанных форм). Каждой семантической форме предъявления ответа соответствует так называемое решающее правило системы «Control», которое оценивает ответы учащегося, предъявленные в указанной форме. В результате для компьютера оказывается возможным игнорировать несущественные ошибки в ответе, учитывать сокращения и синонимы, разрешить, запретить или ограничить перестановки слов, и т.п. В статье приводятся примеры «открытых» тестовых заданий как по естественнонаучным, так и по гуманитарным дисциплинам, варианты возможных правильных и неправильных ответов на них, и соответствующие оценки, выставленные системой «Control». Статистический эксперимент показал, что эти оценки совпадают с оценками, выставленными преподавателями в результате письменного опроса. Включение в сферу компьютерного тестирования «открытых» тестовых заданий требует введения ряда новых понятий для их классификации, которые также обсуждаются в статье.

**Ключевые слова:** аналогия, меры сходства, тестирование знаний, открытые ответы

**ACM Classification Keywords:** I.2.6 Artificial Intelligence – Learning – Analogies; K.3.1 Computers and Education - Computer Uses in Education

---

### Введение

«Внешнее» тестирование знаний учащихся средней и высшей школы призвано решить, в частности, следующие задачи: 1) устранение субъективных факторов в оценке знаний (и, как следствие, коррупции в сфере образования), 2) возможности сравнения по единым критериям как учебных заведений, так и систем образования регионов одной страны и разных стран. Нет сомнения – эти задачи важны. Почему же многие известные ученые и педагоги, прекрасно осознающие эту важность, не покидают лагерь критиков тестирования?

На мой взгляд, если выразить суть предмета разногласий несколькими словами, то ими будут «проблема свободных (открытых) ответов». Какое место в тестировании могут (или должны) занимать вопросы, допускающие ответы в свободной форме? В «традиционных» методиках проверки знаний они преобладают. Но существующие технологии тестирования от них почти отказались: «Для задач, допускающих вариации формы ответа, *правильный ответ должен быть включен в формулировку теста*» — вот тезис, являющийся подлинным предметом дискуссий. Различия вариантов этого тезиса, вроде «Приводится список ответов, среди которых есть как верные, так и неверные» или «Ответ должен быть сконструирован из частей, указанных в тесте, но не объединенных в нем надлежащим образом», не носят принципиального характера.

Ответы в свободной форме «плохи» тем, что их проверка требует значительных затрат при письменном тестировании; и, как обычно считается, невозможна (недостойна) при компьютерном. Цель данной статьи – оспаривание этого «обычного» мнения о недостоверности.

---

### **Выводы по аналогии – базис алгоритмов оценки ответов «открытой» формы**

---

В работах [Леоненко и Поддубный, 1996], [Леоненко, 2002], [Леоненко, 2007] разработана методология оценки ответов свободной формы и необходимый ей математический аппарат. Методы, предложенные в этих работах, суть различные типы *выводов по аналогии* [Уемов, 1971]. Объединяет эти типы то, что во всех случаях признаком, переносимым с модели («правильного ответа») на прототип (фактический ответ), является *качество* фактического ответа, которому соответствует *оценка*, выставленная за ответ.

Я охарактеризую здесь кратко те типы выводов по аналогии, которые используются в системе компьютерного тестирования CONTROL [Леоненко и Поддубный, 1996], [Леоненко, 2009]:

1. *Различные виды парадейгмы*. Парадейгма (см. [Уемов, 1971]) используется, в частности, при сравнении ответов, представляющих собой числовые множества и последовательности (процедуры сравнения здесь исходят из меры пересечения множества-модели и множества-прототипа).

2. *Аналогии через интерпретации формул*. Модель и прототип, представляющие собой математические выражения, сравниваются путем вычисления их значений на одних и тех же значениях аргументов. Разновидности аналогий этого типа зависят от вида сравниваемых выражений (функции, уравнения, тождества) и других факторов. При этом можно разрешить учащемуся «свободно» выбирать обозначения переменных формулы.

3. *Аналогии, базирующиеся на оценке подобия текстов*. Большинство типов таких аналогий, применяемых в системе CONTROL, используют теорию подобия символьных последовательностей, описанную в [Леоненко и Поддубный, 1996], [Леоненко, 2002], [Леоненко, 2007]. Разные виды аналогий этого класса связаны с учетом либо игнорированием весов элементов текста, структуры текста, а также с дополнительными ограничениями.

4. *«Комбинированные» аналогии, применяющие к модели и прототипу несколько разнотипных процедур сравнения*. В частности, такие аналогии используются, когда форма предъявления одинаковых по смыслу прототипов может быть разной (скажем, текст или число); или когда прототип содержит разнородные части (текст и формулу, либо семантически различные фрагменты текста).

Применение выводов по аналогии к оценкам подобия текстов различной природы позволяет системе компьютерного тестирования CONTROL предъявлять студенту вопросы, предполагающие открытую форму ответов, в том числе на естественном языке. CONTROL может *игнорировать несущественные ошибки в словах ответа, учитывать сокращения и синонимы, разрешить, запретить или ограничить перестановки слов*; и т.д. В следующем разделе приведены примеры, иллюстрирующие это.

---

### **Примеры оценок системой CONTROL тестовых заданий с «открытыми» ответами**

---

Следующие далее примеры тестовых задач относятся к различным учебным дисциплинам средней школы и вуза. Они представлены на различных этнических языках, во-первых, с целью показать применимость упомянутых выше методов к этим языкам, и во-вторых, поскольку часть вопросов и ответов взята из учебников и сборников тестов, я счел неуместным переводить их.

Для каждого тестового вопроса я приведу несколько возможных правильных либо приемлемых ответов (буду их нумеровать G1, G2 и т.д. – от «Good»), и несколько неправильных ответов (нумерую их B1, B2 и т.д. – от «Bad»). Система CONTROL адекватно распознаёт эти ответы. В случаях, когда ответ является приемлемым, но не «отличным», ниже приводится оценка, выставленная за него.

Вопросы и возможные ответы на них набраны курсивом, мои комментарии – прямым шрифтом. Первые два вопроса взяты из школьного учебника по информатике [Жалдак і Морзе, 2000].

**Вопрос1.** *Від чого залежить набір послуг, які можна отримати, використовуючи комп'ютер?*

**G1.** *Від набору фізичних пристроїв, що приєднані до комп'ютера; а також від наявних на ньому програм та файлів даних (інформаційного забезпечення).* Это «эталонный» ответ из учебника [Жалдак і Морзе, 2000].

**G2.** Від програмного та апаратного забезпечення.

**G3.** Від техніки та тих програм, що є на цьому комп'ютері.

**V1.** Від програм, що є на цьому комп'ютері. (Предполагаем, как это имеет место в цитируемом учебнике, что в правильном ответе упоминание аппаратного обеспечения является обязательным).

**V2.** Від моделі комп'ютера.

**Вопрос2.** Що називають файлом?

**G1.** Запис (сукупність даних), розміщений на деякому носії, який може бути прочитаний чи переписаний на інший носій окремо (незалежно) від інших записів (ответ из учебника [Жалдак і Морзе, 2000]).

**G2.** Поіменовану купність даних. (Грамматически правильно было бы: «Поіменовану сукупність даних». В каждом слове ответа допущены ошибки. Но такой ответ на вопрос по информатике следует признать правильным).

**V1.** Дані на магнітному диску.

**V2.** Запис у каталозі.

**Вопрос3.** Чему равно  $(a+b)^2$  ?

**G1.**  $a^2+2ab+b^2$

**G2.**  $b^2+aa+2ab$

**V1.**  $(a+b)(a+b)$

**V2.**  $(a+b)^2$ .

**Вопрос4.** Чему равно отношение **длины** окружности к ее **диаметру**?

**G1.** Оно равно числу  $\pi$

**G2.**  $Pi$

**G3.** 3.14

**V1.** Система CONTROL сочтет неправильным любой ответ, не содержащий принятых обозначений числа  $\pi$  или его численного значения (с некоторой приемлемой точностью).

**Вопрос5.** Введите формулу, определяющую **объем** прямого кругового конуса с высотой **H** и радиусом основания **R**.

**G1.** CONTROL сочтет правильной любую формулу, эквивалентную  $\pi R^2 H / 3$ .

**V1.** Что угодно, не эквивалентное формуле  $\pi R^2 H / 3$ .

**Вопрос6.** What the RHOMBUS is? Please, give the verbal definition.

**G1.** Rhombus is a parallelogram with equal sides.

**G2.** It is an equilateral rectangle.

**V1.** A parallelogram with different sides.

**Вопрос7.** З якими історичними подіями періоду Київської Русі пов'язана дата: **988 рік**?

**G1.** Хрещення Русі.

**G2.** Крещение Руси. (Этот ответ дан на другом языке, чем вопрос. Разумеется, нельзя требовать от системы тестирования, чтобы она распознавала имеющие одинаковый смысл ответы, представленные на любых языках. Но в данном случае это возможно, благодаря близости русского и украинского).

**G3.** Принятие Киевской Русью христианской религии.

**V1.** Любой ответ, в котором не упомянуто христианство или Русь.

**Вопрос8.** Де і коли відбулась перша велика битва козаків під проводом Богдана Хмельницького з польським військом?

**G1.** Під Жовтими Водами 5-6 травня 1648 року. («Эталонный» ответ из сборника тестов по истории Украины; CONTROL оценивает его на «отлично»).

**G2.** Жовті Води, травень 1648. (CONTROL оценит этот ответ на «хорошо»).

**G3.** 1648, *Жовті Води*. («Удовлетворительно»).

**B1.** Під Корсунем у 1648 році.

**Вопрос9.** Введіть формулу для вичислення **производной** сложной функции  $F(1/g(x))$  (используйте апостроф для обозначения производных:  $F'$ ,  $g'$  и т.п.).

**G1.** Любая формула, эквивалентная формуле  $F'(1/g(x)) g'(x)/g^2(x)$ .

**B1.** Что угодно, не эквивалентное формуле  $F'(1/g(x)) g'(x)/g^2(x)$ .

**Вопрос10.** Введіть **общее уравнение прямой на плоскости**  $xOy$ .

**G1.**  $Ax+By+C=0$

**G2.**  $ux+vy+w=0$

**G3.**  $C+yB= -Ax$

**B1.**  $y=kx+b$

**B2.**  $Ax+By=0$ .

**Вопрос11.** Что такое указатель (**pointer**)? Дайте определение этого понятия.

**G1.** Переменная, значение которой — адрес области памяти.

**G2.** Это величина, равная адресу какой-то другой величины.

**G3.** Это тип данных, представляющих собой адреса.

**B1.** Это переменная, объявленная с помощью  $\wedge$ .

**B2.** Это адрес оперативной памяти.

**Вопрос12.** Какую задачу выполняет следующий фрагмент программы?

```

...
var S : string; k : byte;
begin
S:=Edit1.Text;
k:=Length(S);
while ( k>0 ) and ( S[1]=' ' ) do
begin DELETE( S, 1, 1); k:=k-1 end;
Edit1.Text:=S;
...

```

**G1.** Вычеркиваются все пробелы в начале строки.

**G2.** Удаление ведущих пробелов.

**G3.** Уничтожаем крайние левые пробелы в строке.

**B1.** Любой ответ, не содержащий упоминания об удалении пробелов, а также о расположении оных в строке.

**Вопрос13.** Запишіть формулу, що виражає закон де Моргана для кон'юнкції.

**G1.**  $\neg(P\&Q) \leftrightarrow (\neg P \vee \neg Q)$

**G2.**  $\neg(a\&b) \leftrightarrow \neg a \vee \neg b$

**G3.**  $(\neg(x\&y) \rightarrow (\neg x \vee \neg y)) \& ((\neg x \vee \neg y) \rightarrow \neg(x\&y))$

Сравнивая эти правильные ответы, вы видите, что студент может произвольно выбирать переменные, входящие в искомую тавтологию. Ясно, что указание вроде «в ответе используйте переменные P и Q» снизило бы качество вопроса.

**B1.**  $\neg(P\&Q) \leftrightarrow (\neg P \& \neg Q)$

**B2.**  $\neg(P\&Q) \leftrightarrow \neg(P\&Q)$ .

**Вопрос14.** Що називають тавтологією у логіці висловлювань? Дайте визначення і наведіть приклад тавтології.

**G1.** Формула, що має значення "істина" незалежно від значень істинності своїх підформул; наприклад  $P \rightarrow P$ . (Оценка CONTROL: «отлично»).

**G2.** Висловлювання, таблиця істинності якого містить лише одиниці; наприклад  $P \rightarrow (P \vee Q)$ . («Отлично»).

**G3.** Формула, таблиця істинності якої містить лише одиниці; наприклад  $(P \vee Q) \rightarrow P$ . («Удовлетворительно», формула не являється тавтологією).

**G4.** Формула, що завжди має значення "істина"; наприклад  $P \rightarrow P$ . («Удовлетворительно», определение неточное).

**B1.** Формула, що завжди має значення "істина"; наприклад  $(P \vee Q) \rightarrow P$ . («Неудовлетворительно»).

**B2.** Логічну помилку (повторення того самого), наприклад "Масло масляне". («Неудовлетворительно», ведь вопрос касался исчисления высказываний, где термин "тавтология" имеет иной смысл).

**Вопрос15.** Запишіть формулы Виета для уравнения:

$$ax^2+bx+c=0 \quad (a \neq 0)$$

и объясните **смысл** величин, входящих в эти формулы.

**G1.** Чтобы CONTROL оценила ответ на «отлично», он должен включать:

a) любые два равенства, эквивалентные равенствам

$$x_1+x_2 = -b/a \quad \text{и} \quad x_1x_2 = c/a;$$

b) указание, что  $x_1$  и  $x_2$  – корни исходного уравнения.

При этом вместо  $x_1$  и  $x_2$  разрешено использовать другие обозначения для корней:  $p$  и  $q$ , и.т.п.

**G2.** Ответ, в котором отсутствует пункт b, но верны оба равенства, CONTROL оценит как «хороший».

**G3.** Ответ, в котором присутствует пункт b, но неверно одно из равенств, будет оценен как удовлетворительный.

**B1.** Любые иные ответы будут сочтены неудовлетворительными.

### О типологии «открытых» тестовых заданий и ответов на них

Приведенные выше примеры показывают, что для компьютерного тестирования оказываются допустимыми (в частности) вопросы следующих – общепринятых в педагогической практике по любой предметной дисциплине – типов:

1. «Свободно» сформулировать описание некоторого факта или ситуации (или же «свободно» дополнить, завершить предложенное описание).
2. Сформулировать определение некоторого понятия.
3. Дать объяснение некоторому факту или ситуации.
4. Привести примеры, иллюстрирующие некоторое понятие или ситуацию.
5. Привести контр-примеры к чему-либо.
6. Описать границы или условия, необходимые для чего-либо.
7. Исправить ошибки в предьявленном определении, описании или выражении.
8. Дать иную, равносильную формулировку предложенного в задании определения, утверждения, формулы.
9. Выразить символически некоторое утверждение, сформулированное на естественном языке (или, наоборот, дать словесное истолкование некоторой формулы).

Включение таких «открытых» тестовых заданий в число подлежащих автоматической проверке требует уточнения ряда понятий, связанных с компьютерным тестированием. Ниже обсуждаются некоторые из них.

Назовем темой вопроса множество тех понятий и связей между ними, которые должны быть упомянуты в правильном и полном ответе на данный вопрос.

Назовем тематической формой вопроса структуру, которая связывает понятия из темы вопроса в данной формулировке вопроса.

Используя эту терминологию, можно сказать, что приведенные выше типы 1 – 9 суть примеры некоторых видов *тематических форм* вопроса (поскольку ясно, что вопросы, различающиеся по виду 1 – 9, могут иметь одну и ту же тему вопроса).

Назовем формой изображения вопроса структуру, в которой видео- или аудио-изображения понятий и связей темы вопроса предъясняются учащемуся.

Традиционно применяемые «закрытые» формы тестовых заданий («выбрать верный ответ», «дополнить», «установить соответствие» и др.), как представляется, являются видами форм изображения вопроса.

Каждый опытный преподаватель, по-видимому, будет согласен с тем, что *выбор тематической формы вопроса ограничивает множество пригодных для неё форм изображения вопроса*. Иными словами, не любая форма изображения вопроса пригодна для заданной его тематической формы. Например, не просто совместить упомянутые выше тематические формы 1 – 9 с формой изображения «выбрать верный ответ» без снижения «качества» вопроса.

Назовем семантической формой предъяснения ответа на вопрос структуру изображения понятий из темы вопроса в ответе, которую данный вопрос предписывает (в явном или неявном виде) учащемуся.

Имеется тесная связь между формой изображения вопроса и семантической формой предъяснения ответа на него. В большинстве случаев учащийся узнает, *как именно ему формулировать ответ, именно из поставленного вопроса (или из контекста, его сопровождающего)*. Поэтому тематическая форма вопроса оказывает – через «навязываемое» ею ограниченное множество форм изображения вопроса – значительное, если не решающее, влияние на семантическую форму предъяснения ответа на данный вопрос.

В частности, вопрос, поставленный в какой-нибудь из перечисленных выше тематических форм 1 – 9, предполагает соответствующую семантическую форму предъяснения ответа. Так, ответом на вопрос формы 4 *не должно быть определение* понятия (если, конечно, оно не сопровождается *примером*); а для вопроса формы 2 все обстоит наоборот.

Придерживаясь традиционной терминологии, следовало бы сказать, что вопросы видов 1 – 9 предполагают *открытую* форму предъяснения ответа, поскольку не предписывают учащемуся ограничений на его формулировку. Однако при этом, очевидно, *не предполагается, что учащийся свободен в выборе семантической формы предъяснения своего ответа*. То, что имеется здесь в виду под «открытостью», означает свободу выбора учащимся средств аудио- или видео-изображений понятий, предполагаемых темой вопроса – но не выбора самих понятий и той их взаимосвязи, которая «навязана» тематической формой вопроса.

Назовем синтаксической формой предъяснения ответа на вопрос избранные учащимся аудио- или видео-средства изображения, а также структуру взаимосвязи аудио- или видео-элементов в предъясненном учащимся ответе.

Таким образом, в данной терминологии следует сказать, что тематические формы вопроса 1 – 9 предполагают *свободную синтаксическую форму* ответа на них.

Систему компьютерного тестирования знаний «Control» можно рассматривать как «библиотеку», предоставляющую преподавателю *средства оценки ответов* предъясненных в разнообразных *семантических формах* (причем эта библиотека может пополняться новыми типами указанных форм). Каждой семантической форме предъяснения ответа соответствует так называемое *решающее правило*



системы «Control», которое оценивает ответы учащегося, предъявленные в указанной форме [Баранов и Леоненко, 2004].

При этом большинство решающих правил *не накладывает никаких ограничений на синтаксические формы* предъявления соответствующих ответов (или стремится свести такие ограничения к минимуму).

---

### Заключение

За несколько лет применения системы «Control» в Одесской национальной академии связи им. А.С.Попова разработаны банки задач по курсам "Информатика", "Аналитическая геометрия на плоскости", "Дифференциальное исчисление", "Численные методы решения инженерных задач", "Теория линейных электрических цепей", "Математическая логика".

Были выполнены статистические оценки эффективности системы CONTROL. В эксперименте участвовали все студенты первого курса Академии (более 500 человек). Половина сдавала регулярные «зачеты» системе CONTROL, а половина писала в то же время письменные контрольные работы. Задачи тестирования в обеих группах совпадали.

Эксперимент показал [Баранов, 2004], что итоговые уровни качества знаний студентов в обеих группах статистически совпали. Это позволяет утверждать, что компьютерное тестирование может использоваться в учебном процессе с не меньшей эффективностью, чем традиционные методы контроля знаний.

---

### Библиография

- [Leonenko, 2002] L. Leonenko. Analogical inferences in computer assisted knowledge testing systems // 6th Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2002), Orlando, Florida, USA, 2002. Proceedings, Vol. XVIII, p.371-376.
- [Баранов, 2004] В. Ю. Баранов. Комп'ютерне тестування з інформатики: підсумки педагогічного експерименту в Одеській національній академії зв'язку // Теорія та методика навчання математики, фізики, інформатики. – Вип. 4. – Т. 3 – Кр.Ріг.: НМетАУ, 2004. – с. 6 –12.
- [Баранов и Леоненко, 2004] В. Ю. Баранов, Л. Л. Леоненко. Компьютерное тестирование знаний: об алгоритмах оценки «свободных» ответов. // «ИНФОТЕХ-2004». Материалы междунаrod. науч.-практич. конф. Киев-Севастополь: НТО РЭС Украины, 2004, с. 163 – 171.
- [Жалдак і Морзе, 2000] М. І. Жалдак, . Н. В Морзе . Інформатика-7. Експериментальний навчальний посібник для учнів 7 класу загальноосвітньої школи. – К.: ДіаСофт, 2000. – 208 с.
- [Леоненко, 2007] Л. Л. Леоненко. Алгоритмы оценки аналогичности текстов и их применение в компьютерном тестировании // Сб. трудов VII международной конф. "Интеллектуальный анализ информации". – К.: Просвіта, 2007, с. 210–220.
- [Леоненко, 2009] Л. Л. Леоненко. Оценивание «открытых» ответов в системе компьютерного тестирования знаний CONTROL // Сб. трудов IX международной конф. "Интеллектуальный анализ информации". – К.: Просвіта, 2009, с. 240–246.
- [Леоненко и Поддубный, 1996] Л.Л. Леоненко, Г.В. Поддубный. Теория подобия конечных последовательностей и ее приложения к распознаванию образов // Автоматика и телемеханика, 1996, № 8, с.119-131.
- \*[Уемов, 1971] А.И. Уемов. Логические основы метода моделирования. М.: Мысль, 1971. – 311 с.

---

### Информация об авторе



**Леонид Леоненко** – Одесская национальная академия связи им. А. С. Попова, доцент, ул. Конная, 22, кв. 6, Одесса, 65029, Украина; e-mail: Leonid.Leonenko@gmail.com  
Основные области научной деятельности: неклассическая логика, общая теория систем, компьютерное тестирование знаний

---

## Mathematical Foundations

---

### ЕВКЛИДОВЫ ПРОСТРАНСТВА: КОНСТРУКТИВНЫЕ МЕТОДЫ ОПИСАНИЯ БАЗОВЫХ СТРУКТУР И ИХ ИСПОЛЬЗОВАНИЕ

Владимир Донченко

**Аннотация:** Рассмотрены базовые структуры стандартных типов евклидовых пространств: числовых векторов и матриц фиксированных размерностей. Под базовыми имеются в виду линейные и нелинейные объекты, принципиальные с точки зрения их использования в прикладных задачах. Развита конструктивная методика построения базовых структур, их описания, взаимного перехода и использования на основе систематического развития и применения аппарата сингулярного представления, а также псевдообращения по Муру - Пенроузу. Теория сингулярного разложения и псевдообращения по Муру-Пенроузу распространена на матрицы, как линейные отображения между евклидовыми пространствами матриц. Приведены и доказаны соответствующие теоремы. Развита теория группирующих операторов, важная в прикладном отношении как инструмент выявления и использования групповых свойств объектов в евклидовых пространствах. Кроме того, группирующие операторы позволяют, в частности, прояснить алгебраическую суть расстояния Махаланобиса.

**Ключевые слова:** Псевдообращение по Муру - Пенроузу, сингулярное представление матрицы как матричного отображения, псевдообращение матричного отображения, эллипсы группировки, расстояния соответствия в кластеризации, расстояние Махаланобиса.

**ACM Classification Keywords:** G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.

---

#### Вступление

В работе предложена и обоснована концепция базовых структур евклидова пространства, к которым предлагается отнести структуры двух основных видов: линейные и нелинейные. И те и другие структуры евклидова пространства проявляются либо во множественной форме: подпространства и гиперплоскости минимальные эллипсы группировки, либо - в единичной: линейные операторы, квадратичные формы. Отметим, что к базовым нелинейным структурам отнесены, квадратичные формы (в работе – неотрицательно определённые) и «характерные» множества на их основе: эллипсы группировки, а также минимальные эллипсы группировки: поверхности уровня отвечающие единичному значению соответствующих квадратичных форм. Примечательной особенностью эллипсов группировки является то, что они содержат все векторы набора векторов, и что такие эллипсы можно выбрать «минимальными» для них. Описаны конструктивные методы генерации базовых структур в связи с порождающей их совокупностью векторов, а также формулы взаимного перехода от одних типов структур к другим: от линейных подпространств и гиперплоскостей к матрицам и наоборот, а также от набора векторов к матрицам квадратичных форм и эллипсам группировки. В числе других рассмотрены конструктивные способы порождения подпространств и гиперплоскостей, а также ортогональных проекторов, связанных с указанными объектами. В том же русле конструктивности лежит рассмотрение концепции группирующих операторов и минимальных эллипсов группировки. Концепция группирующих операторов позволяет использовать в прикладных исследованиях связь того или иного набора векторов с важным видом нелинейных структур евклидовых пространств: минимальными эллипсоидами группировки. И в линейном,

и в нелинейном случае упомянутая конструктивность обеспечивается применением псевдообращения по Муру – Пенроузу (ПДО), а также новыми результатами в этой области, берущими своё начало и опирающимися на фундаментальную работу Н.Ф Кириченко [Кириченко, 1997]. Для расширения возможности аппарата, предложено развитие теории псевдообращения на матричные отображения между евклидовыми пространствами матриц фиксированных размерностей со следовыми скалярными произведениями. Кроме того, на случай отображений между евклидовыми матричными пространствами перенесена теорема о сингулярном разложении матрицы оператора, доказана теорема свёртки о соотношении между сингулярными разложениями одной и той же матрицы в векторном и матричном случаях.

Важность и эффективность применения приведённых результатов определяется возможностью их использования для решения широкого класса прикладных задач, включая линейную регрессию, в том числе векторную, теорию оптимального управления, кластеризацию, прогноз и функциональные сети, являющиеся обобщением искусственных нейронных сетей.

Как отмечалось в работе [Донченко,2009], «структура объекта» ассоциируется с реализацией представления об объекте, как чём-то едином, составленном из взаимодействующих, связанных между собою частей. Математическое описание, модель объекта, реализует описание «структуры» объекта средствами математического описания «связей». Это означает, что структура объекта должна быть передана, отражена в математической модели, средствами математического структурирования. К последним относятся четыре базовых структуры, а также их комбинации. К упомянутым четырём базовым структурам математики относятся: отношения, функции, операции и наборы подмножеств. К комбинациям относятся, например, структуры линейного и евклидова пространств.

Структура линейного пространства, занимает ведущее место среди важнейших математических структур. В свою очередь, к базовым структурам таких пространств можно отнести линейные и нелинейные объекты, являющиеся принципиальными с точки зрения использования. Это, с одной стороны объекты на основе матриц: линейные операторы и квадратичные формы, с другой - структуры, которые можно назвать множественными. К ним можно отнести подпространства и гиперплоскости, а также эллипсоиды, порождаемые неотрицательно определёнными квадратичными формами.

Важную роль в прикладных исследованиях играют конструктивные методы описания соответствующих объектов. Принципиальным в этом отношении применительно к базовым линейным структурам является использование ортогональных проекторов. Для нелинейных - использование группирующих операторов. И в том и другом случае аппарат псевдообращения позволяет конструктивно строить как те, так и другие. Именно применение этого аппарата является основой для решения основных задач, связанных с построением, описанием и использованием базовых структур линейных, точнее, - евклидовых пространств.

Отметим, что ортогональные проекторы играют важную роль в исчерпывающем исследовании систем линейных алгебраических уравнений (СЛАУ). Принципиально важны они также и в постановке и решении важных оптимизационных задач с квадратичными функционалами качества в евклидовых пространствах. Важная роль ортогональных проекторов в решении прикладных задач в полной мере сохраняется и в связи с задачей построения наилучших квадратичных приближений правой части СЛАУ значениями левой, когда СЛАУ не имеет точных решений. Такие наилучшие приближения называют также псевдорешениями. Конструктивное описание ортогональных проекторов в связи с естественными подпространствами линейного оператора прямо определяется псевдообращением по Муру – Пенроузу [Moore,1920],[Penrose, 1955] (см. также, например,[Алберт, 1977]).

Следует отметить, также, что важную роль в применении аппарата псевдообращения играет сингулярное представление (его называют также SVD - представлением) произвольной  $m \times n$  матрицы в специфической форме записи в виде взвешенной суммы тензорных произведений специального набора пар векторов с прозрачным геометрическим содержанием. При этом матрица рассматривается как

матрица линейного оператора между двумя пространствами числовых векторов  $R^n$  и  $R^m$ . Это определяет специфику сингулярного разложения и интерпретацию его составляющих.

Заметим, однако, что одна и та же  $m \times n$  матрица  $A$  задаёт как линейный оператор  $A: R^n \rightarrow R^m$  между евклидовыми пространствами числовых векторов, так и линейный оператор  $A: R^{n \times p} \rightarrow R^{m \times p}$  для произвольного натурального  $p$ . Переход к евклидовым пространствам  $R^{n \times p}, R^{m \times p}$  матриц со следовыми скалярными произведениями требует учёта специфики этих евклидовых пространств для передачи принципиальных особенностей SVD-представления. В работе приведено SVD-представление для матриц, рассматриваемых как линейные операторы между пространствами матриц, доказана теорема свёртки, связывающие два варианта SVD-представления матрицы: векторного и матричного.

Важным в описании нелинейных структур являются минимальные эллипсы группировки и, как и в линейном случае, конструктивные способы их описания. Говоря о минимальных эллипсах группировки, будем иметь в виду «минимальный эллипс», включающий заданный набор векторов. Как оказывается, матрицей квадратичной формы для такого эллипса является группирующий оператор, конструктивно описываемый средствами псевдообращения. Квадратичная форма, отвечающая группирующему оператору, в искажающем суть такого оператора виде, фигурирует в «статистическом» варианте расстояния Махаланобиса [Mahalanobis, 1936] (см. также, например, [McLachlan, Geoffry, 1992]), когда вместо ковариационной матрицы многомерного нормального распределения используется стандартная её оценка на основе выборки векторов  $a(1), \dots, a(n)$ .

В заключение отметим, что основные идеи, дух и результаты предлагаемой работы восходят к работам и используют результаты развитой в них теории псевдообращения нашего безвременно ушедшего коллеги, друга и учителя, профессора Н.Ф. Кириченко.

### Базовые линейные структуры и связи между ними

В дальнейшем, говоря о евклидовом пространстве  $R^p$ , будем иметь в виду множество конечных числовых последовательностей одной и той же длины  $p$ , записанных в столбик с покомпонентными операциями сложения и умножения на скаляр и суммой покомпонентных произведений в качестве скалярного произведения. Именно такой вариант евклидова пространства будем стандартным образом обозначать через  $R^p$ , а его элементы – через  $a, a^T = (a_1, \dots, a_p)$ . Стандартные ортонормированные базисы, составленные из векторов с единственной единичной компонентой (остальные – нули) на месте с соответствующим номером будут обозначаться для  $R^m$  и  $R^n$  соответственно через  $e(j) \in R^m, j = \overline{1, m}, e_{(i)} \in R^n, i = \overline{1, n}$ , а для общего случая –  $e_k \in R^p, k = \overline{1, p}$ . Оператор  $A$  из  $R^n$  в  $R^m$ :  $A: R^n \rightarrow R^m$ , задаваемый в ортонормированных базисах  $e(j) \in R^m, j = \overline{1, m}, e_{(i)} \in R^n, i = \overline{1, n}$ , будем отождествлять с  $m \times n$  - матрицей  $A = (a_{ij})$  этого оператора в этих базисах. Для матрицы  $A = (a_{ij})$  будем использовать также блочное представление по столбцам (столбцовое) и строкам (строчное):

$$A = \begin{pmatrix} a_{(1)}^T \\ \dots \\ a_{(m)}^T \end{pmatrix} = (a(1) : \dots : a(n)), a_{(i)} \in R^n, i = \overline{1, m}, a(j) \in R^m, j = \overline{1, n}.$$

Пространство всех  $p \times q$  матриц с покомпонентным умножением и сложением, а также «следовым» скалярным произведением будем обозначать  $R^{p \times q}$ . «Следовое» скалярное произведение  $(\cdot, \cdot)_{tr}$  двух матриц определяется соотношением

$$(C, D)_{tr} = \text{tr} C^T D = \sum_{i=1}^q \sum_{k=1}^p c_{ki} d_{ki}, C = (c_{ij}), D = (d_{ij}) \in R^{p \times q}.$$

Линейное подпространство, порождённое системой векторов,  $c_k \in R^p, k = \overline{1, K}$ , будет обозначаться через  $L(c_k, k = \overline{1, K}) \equiv L(c_1, \dots, c_K)$ , а линейное подпространство значений линейного оператора  $A: R^n \rightarrow R^m$  – через  $L_A$ , соответственно, для  $A^T$  – через  $L_{A^T}$ .

Первым из набора свойств, описывающих связи между базовыми структурами, является следующее утверждение.

1. Наборы векторов, подпространства, матрица, составленная из набора векторов. Справедливо следующее утверждение.

$$L_A = L(a(1), \dots, a(n)).$$

Таким образом, линейное подпространство, порождённое набором векторов, совпадает с подпространством значений матрицы, составленной из векторов набора, как из столбцов.

2. «Векторы из матрицы». Для элементов столбцового и строчного представления матрицы  $A \in R^{m \times n}$  справедливы соотношения

$$a(j) = A e_{(j)}, j = \overline{1, n}, a_{(i)}^T = e^T(i) A, i = \overline{1, m}.$$

3. Произведение матриц как сумма тензорных произведений их составляющих. Для произведения произвольных матриц  $B, C$  со столбцовым и строчным представлением

$$B = (b(1): \dots : b(r)), b(j) \in R^m, j = \overline{1, r}, C = \begin{pmatrix} c_{(1)}^T \\ \dots \\ c_{(r)}^T \end{pmatrix}, c_{(r)}, c(i) \in R^n, i = \overline{1, r}$$

соответственно, а также диагональной матрицы  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$  справедливо соотношение

$$B \Lambda C = \sum_{i=1}^r \lambda_i b(i) c_{(i)}^T.$$

Важной составляющей аппарата конструктивного описания и использования линейных структур является понятие ортогонального проектора (в дальнейшем ОП), определение которого полностью согласуется со стандартным геометрическим представлением об ортогональном проектировании. Общей, основой эффективного использования ортогональных проекторов, как и возможности их конструктивного построения в связи с линейными подпространствами, является наличие двух эквивалентных определений таких проекторов, а также возможность их описание через псевдообращение.

4. «Геометрическое» определение ОП. Для разложения  $R^p = L + L^\perp$  в прямую сумму ортогональных подпространств ортогональным проектором  $P_L$  на линейное подпространство  $L \subseteq R^p$  называется оператор, определяемый соотношением  $P_L x = P_L(x_L + x_{L^\perp}) = x_L$ , где

$$x = x_L + x_{L^\perp}, x_L \in L, x_{L^\perp} \in L^\perp \quad (1)$$

– однозначное разложение произвольного вектора  $x \in R^p$  по двум составляющим ортогональной суммы. Очевидным образом оператор ортогонального проектирования является линейным оператором.

5. Дополнительные друг к другу ОП. Разложение (1) произвольного вектора  $x \in R^p$  в силу симметричности относительно ортогональных слагаемых определяет одновременно два ортогональных проектора:  $P_L, P_{L^\perp}$  с очевидным соотношением

$$P_L + P_{L^\perp} = E_p$$

где  $E_p$  – единичная матрица соответствующей размерности.

6. *Дополнительные друг к другу ОП и соответствующие им подпространства.* Для ортогонального проектора  $P_L$  на подпространство  $L$  оператор  $Z_L \equiv E_p - P_L$  является ортогональным проектором на ортогональное дополнение  $L^\perp$ :  $Z_L \equiv E_p - P_L = P_{L^\perp}$ .

7. *Абстрактное определение ОП.* Для того, чтобы линейный оператор  $P: R^p \rightarrow R^p$ , был оператором ортогонального проектирования необходимо и достаточно, чтобы он был идемпотентным и симметричным оператором. Линейное подпространство  $L_p$ , на которое совершается ортогональное проектирование в соответствии с «геометрическим определением» описывается одним из двух соотношений:

$$L_p = \{x: x = Pu, u \in R^p\} = \{x: x = Px, x \in R^p\}.$$

8. *Классическое или векторное сингулярное (SVD-) представление произвольной  $m \times n$  матрицы.* Для произвольной  $A \in R^{m \times n}$  ранга  $r \leq \min(m, n)$ , рассматриваемой как линейное отображение между евклидовыми пространствами векторов, справедливо следующее представление матрицы в виде взвешенной суммы тензорных произведений векторов

$$A = \sum_{i=1}^r \lambda_i u_i v_i^T \quad (2)$$

где

$\lambda_1^2 \geq \dots \geq \lambda_r^2 > 0$  общий набор ненулевых собственных чисел матриц  $AA^T, A^T A$ ;

$u_i \in R^m, i = \overline{1, r}$ , - ортонормированный набор собственных векторов матрицы  $AA^T$ , отвечающих ненулевым собственным числам:  $AA^T u_i = \lambda_i^2 u_i, \lambda_i^2 > 0, i = \overline{1, r}, u_i^T u_j = \delta_{ij}, i \neq j$ ;

$v_i \in R^n, i = \overline{1, r}$ , -- ортонормированный набор собственных векторов матрицы  $A^T A$ , отвечающих ненулевым собственным числам:  $A^T A v_i = \lambda_i^2 v_i, \lambda_i^2 > 0, i = \overline{1, r}, v_i^T v_j = \delta_{ij}, i \neq j$ .

9. *Комментарий к п.8.* Собственно, действие оператора  $A$  на вектор  $x \in R^n$  в SVD-разложении, представляет собой перенос разложения вектора по ортогональному базису или его части  $v_i \in R^n, i = \overline{1, r}$ , с координатами  $x_i = v_i^T x = (v_i, x)_{R^n}, i = \overline{1, r}$ , в одном пространстве в разложение по ортогональному базису  $u_i \in R^m, i = \overline{j, r}$ , или - его части, в другом пространстве с умножением координат на положительные числа, соответственно,  $\lambda_i \in R^m, i = \overline{1, r}$ . Поэтому вариантом записи SVD – разложения (2), может быть следующий:

$$Ax = \sum_{i=1}^r \lambda_i u_i (v_i, x)_{R^n}, x \in R^n.$$

### Матрицы как линейные операторы между матричными евклидовыми пространствами

10. *Матрица  $A \in R^{m \times n}$  как линейный оператор из матричного евклидова пространства  $R^{n \times p}$  в  $R^{m \times p}$ .* Произвольная  $m \times n$  матрица  $A$  может рассматриваться как матрица линейного оператора между двумя евклидовыми пространствами матриц  $R^{n \times p}$  и  $R^{m \times p}$  со следовым скалярным произведением. Этот

оператор описывается стандартным образом: для произвольной матрицы  $X \in R^{n \times p}$  результатом действия оператора является матрица  $AX \in R^{m \times p}$ .

11. *Матричное сингулярное разложение  $m \times n$  матрицы  $A \in R^{m \times n}$  (M-SVD).*

Теорема 1. Справедливо следующее представление матрицы в виде суммы, реализующий перенос разложения по ортонормированным собственным элементам (собственным матрицам) матрицы  $A^T A$  на ортонормированные собственные элементы матрицы  $A^T A$

$$AX = \sum_{i=1}^r \sum_{k=1}^p \lambda_i u_i e_k^T (v_i e_k^T, X)_{tr} \quad (3)$$

где

$\lambda_1^2 \geq \dots \geq \lambda_r^2 > 0$  - общий набор ненулевых собственных чисел матриц  $AA^T, A^T A$ ;

- $u_i e_k^T \in R^{m \times p}, i = \overline{1, r}, k = \overline{1, p}$ , - ортонормированный набор собственных  $m \times p$  матриц матрицы  $AA^T$ , отвечающих ненулевым собственным числам:  $AA^T u_i e_k^T = \lambda_i^2 u_i e_k^T, \lambda_i^2 > 0, i = \overline{1, r}, k = \overline{1, p}, (u_i e_k^T, u_j e_l^T)_{tr} = \delta_{ij} \delta_{kl}, i \neq j, k \neq l$ ;
- $v_i e_k^T \in R^{n \times p}, i = \overline{1, r}, k = \overline{1, p}$ , - ортонормированный набор собственных  $n \times p$  матриц матрицы  $A^T A$ , отвечающих ненулевым собственным числам:  $A^T A v_i e_k^T = \lambda_i^2 v_i e_k^T, \lambda_i^2 > 0, i = \overline{1, r}, k = \overline{1, p}, (v_i e_k^T, v_j e_l^T)_{tr} = \delta_{ij} \delta_{kl}, i \neq j, k \neq l$ .

12. *Совпадение векторного и матричного варианта SVD: теорема свёртки.*

Теорема 2. Векторный и матричный вариант SVD-разложения одной и той же  $m \times n$  матрицы  $A \in R^{m \times n}$  совпадают между собой: для произвольной матрицы  $X \in R^{n \times p}$

$$\left( \sum_{i=1}^r \lambda_i u_i v_i^T \right) X = \sum_{i=1}^r \sum_{k=1}^p \lambda_i u_i e_k^T (v_i e_k^T, X)_{tr} \quad (4)$$

Принципиальную роль в описании базовых структур евклидовых пространств играет псевдообращение по Муру – Пенроузу [Moore, 1920], [Penrose, 1955] как одноместной операции  $A^+$  над прямоугольными матрицами произвольной размерности. В дальнейшем термин «псевдообращение» будем сокращать до ПДО.

С учётом п.10 необходимо различать ПДО для матрицы, как векторного отображения, и ПДО для той же матрицы, как матричного отображения. Первое, собственно, классическое по Муру-Пенроузу, будем обозначать как ВПДО, второе – как МПДО.

---

### Псевдообращение: векторный и матричный вариант отображения

---

13. *Определение ВПДО через векторное SVD - представление матрицы.* Для произвольной  $m \times n$  - матрицы  $A$  её ВПДО  $A^+$  (классическое по Муру-Пенроузу) определяется соотношением

$$A^+ = \sum_{i=1}^r \lambda_i^{-1} v_i u_i^T : R^m \rightarrow R^n \quad (5)$$

14. *Определение МПДО через матричное SVD - представление матрицы.*

Теорема 3. Для произвольной  $m \times n$  - матрицы  $A$  её МПДО  $A_M^+$  определяется соотношением

$$A_M^+ X = \sum_{i=1}^r \sum_{k=1}^p \lambda_i^{-1} v_i e_k^T (u_i e_k^T, X)_{tr} : R^{m \times p} \rightarrow R^{n \times p} \quad (6)$$

15. Совпадение векторного (классического) ПДО и ПДО матричного.

Теорема 4. Матричный и векторный вариант ПДО для произвольной  $m \times n$  - матрицы  $A$  совпадают между собой: т.е. совпадают операторы  $A^+, A_M^+$ , определяемые соотношениями (5), (6) соответственно:

$$A_M^+ = A^+.$$

*Доказательство.* Результат очевидным образом вытекает из теоремы свёртки п.12.

### Евклидовы пространства, базовые линейные структуры и ПДО

16. Основные ортогональные проекторы (ОП), связанные с матрицей линейного оператора. Основными ОП, связанными с матрицей  $A$  являются ОП  $P(A^T), P(A)$  на подпространства  $L_A, L_{A^T}$  соответственно, а также  $Z(A^T), Z(A)$  соответственно - на их ортогональные дополнения в  $R^m, R^n$  соответственно.. Они определяются соотношениями :

$$P(A^T) = AA^+ = \sum_{i=1}^r u_i u_i^T, P(A) = P((A^T)^T) = A^T (A^T)^+ = A^+ A = \sum_{i=1}^r v_i v_i^T,$$

$$Z(A) = E_n - P(A) = E_n - A^+ A, Z(A^T) = E_m - P(A^T) = E_m - A^T A^+ = E_m - AA^+.$$

Важность последних соотношений определяется тем, что  $L_{A^T}^\perp$  является множеством нулей оператора  $A$ .

17. Множество нулей (ядро) оператора  $A$ . Подпространство  $L_{A^T}^\perp$  является ядром,  $Ker A$ , (множеством нулей) оператора  $A$ :

$$L_{A^T}^\perp = Ker A = Z(A)R^n.$$

18. Условие совместности системы линейных алгебраических уравнений (СЛАУ). Для совместности СЛАУ  $Ax = y, y \neq 0$ , необходимо и достаточно, чтобы  $y^T Z(A^T)y = 0$ . В этом случае  $A^+y$  является наименьшим по норме решением. Оно ортогонально к  $Ker A$ , а множество всех решений  $\Omega_y$  описывается соотношением

$$\Omega_y = A^+y + Z(A)R^n = \{x : x = A^+y + Z(A)v, v \in R^n\} \quad (7)$$

19. Псевдорешения СЛАУ. Если СЛАУ  $Ax = y$  несовместна, т. е.  $y^T Z(A^T)y > 0$ , то множество, определяемое соотношением (7) описывает совокупность всех наилучших квадратичных приближений правой части значениями левой:

$$\Omega_y = A^+y + Z(A)R^n = Arg \min_{x \in R^n} \|Ax - y\|^2 \quad (8)$$

Значение невязки для любого наилучшего квадратичного приближения составляет  $y^T Z(A^T)y$ .

20. Для того, чтобы матричное уравнение (МЛУ)  $AX = Y, A \in R^{m \times n}, X \in R^{n \times p}, Y \in R^{m \times p}$ , имело решения необходимо и достаточно, чтобы  $tr Y^T Z(A^T)Y = 0$ . В этом случае множество  $\Omega_Y$  определяется соотношением

$$\Omega_Y = \{X : X = A^+Y + Z(A)V, V \in R^{n \times p}\} \quad (9)$$

В случае, когда МЛУ несовместно, т.е., когда  $tr Y^T Z(A^T)Y > 0$ , множество (9) описывает решение оптимизационной задачи наилучшего квадратичного приближения правой части  $Y$  значениями левой части  $AX$  того же уравнения:



$$\mathop{\text{Arg}} \min_{X \in R^{n \times p}} \|AX - Y\|_{tr}^2 = \Omega_Y = \{X : X = A^+Y + Z(A)V, V \in R^{n \times p}\}. \quad (10)$$

Величина невязки для каждого решения оптимизационной задачи составляет  $\text{tr}Y^T Z(A^T)Y$ .

*Замечание 1.* Обратим внимание, что множество решений МаЛУ, и множество псевдорешений, когда МаЛУ несовместно, описываются одним и тем же соотношением (9).

*Замечание 2.* Поскольку при совместности МаЛУ  $\min_{X \in R^{n \times p}} \|AX - Y\|_{tr}^2 = 0$  и этот минимум достигается на решениях МаЛУ, то множество  $\Omega_Y$  из (9) описывает и множество решений и множество псевдорешений как решение одной и той же оптимизационной задачи

$$\Omega_Y = \{X : X = A^+Y + Z(A)V, V \in R^{n \times p}\} = \mathop{\text{Arg}} \min_{X \in R^{n \times p}} \|AX - Y\|_{tr}^2 \quad (11)$$

*Замечание 3.* Обратим внимание, что матрица  $A^+Y$  является решением, точным или псевдо-, МаЛУ. Это решение будем называть базовым.

21. Оптимизационное свойство Пенроуза для матричного ПДО.

Теорема 5. Базовое решение МаЛУ как в точном, так и в псевдоварианте, является единственным наименьшим по норме решением оптимизационной задачи поиска наилучшего квадратичного приближения правой части МаЛУ значениями левой части того же уравнения:

$$A^+Y = \mathop{\text{Arg}} \min_{X \in \mathop{\text{Arg}} \min_{X \in R^{n \times p}} \|AX - Y\|_{tr}^2} \|X\|_{tr}. \quad (12)$$

*Доказательство.* В силу замечания 2 и множество решений, и множество псевдорешений являются решением одной и той же оптимизационной задачи. Поэтому для любого  $X \in \mathop{\text{Arg}} \min_{X \in R^n} \|AX - Y\|_{tr}^2$  в соответствии с (10) справедливо представление

$$X = A^+Y + Z(A)V, V \in R^{n \times n}. \quad (13)$$

Поскольку  $Z(A)V \in Z(A)R^{n \times p}$  и, как нетрудно убедиться, для любого  $V \in R^{n \times p}$

$$A^+Y \perp_{tr} Z(A)V,$$

то в соответствии с многомерным вариантом теоремы Пифагора

$$\|X\|^2 = \|A^+Y + Z(A)V\|^2 = \|A^+Y\|^2 + \|Z(A)V\|^2 \geq \|A^+Y\|^2.$$

Таким образом, для произвольного  $X \in \mathop{\text{Arg}} \min_{X \in R^{n \times p}} \|AX - Y\|_{tr}$

$$\|X\|^2 \geq \|A^+Y\|^2,$$

и базовое решение  $A^+Y$ , псевдо- или точное, является наименьшим по норме решением оптимизационной задачи, определяемой правой частью соотношения (12). Единственность решения оптимизационной задачи вытекает из того, что  $\|Z(A)V\|^2 = 0 \Leftrightarrow Z(A)V = 0$ .

22. *Линейная зависимость вектора от набора векторов.* Для линейной зависимости вектора  $d \in R^m$  от векторов набора  $a(j) \in R^m, j = \overline{1, n}$ , необходимо и достаточно, чтобы выполнялось соотношение  $d^T Z(A^T)d = 0$  с  $A = (a(1) : \dots : a(n))$ . Очевидным образом, то же условие является необходимым и достаточным для линейной зависимости вектора  $d \in R^m$  от столбцов  $m \times n$  матрицы  $A$ . Отметим также,

что необходимым и достаточным условием линейной зависимости строки  $a^T, a \in R^n$  от строк матрицы  $A \in R^{m \times n}$  является условие  $a^T Z(A)a = 0$ .

23. В связи с ограниченностью работы только упомянем другие важные в приложениях результаты, которые приведены в работе [Донченко et al, 2010]. Это прямые и обратные формулы Гревилля, а также формулы аналитического возмущения ПдО Н.Ф.Кириченко [Кириченко,1997]. В сущности, в этих результатах речь идёт о формулах, связывающих ПдО изменённой (возмущённой) матрицы с ПдО исходной, а также – с характеристиками возмущения. В формулах Гревилля таким возмущением является добавление (прямые) или вычёркивание (обратные) строки или столбца матрицы. Собственно, Гревиллю принадлежит формула, касающаяся добавления строки, да ещё и только для случая независимости добавляемой строки от строк исходной матрицы. Результаты, связанные с добавлением зависимой строки, как и с вычёркиванием произвольного типа строки, принадлежат Н.Ф.Кириченко [Кириченко,1997].

В аналитических формулах возмущения [Кириченко,1997] изменение исходной матрицы происходит аддитивно: через добавление «простейшей» матрицы. В качестве «простейшей» матрицы выступает тензорное произведение  $ab^T$  двух векторов  $a \in R^m, b \in R^n$ . Как и в случае формул Гревилля, вид ПдО возмущённой матрицы определяется тем, являются ли компоненты возмущения зависимыми, или независимыми, от, соответственно, столбцов и строк возмущаемой матрицы. Кроме того, результат зависит также и от того, падает или сохраняется ранг возмущённой матрицы, когда составляющие элементы возмущения зависят от соответствующих составляющих возмущаемой матрицы. Обратим внимание, что все формулы или условия описываются явными аналитическими выражениями. Условия линейной зависимости или независимости для столбцов уже нашли своё отражение в п.22. Для описания зависимости от строк в соответствующем условии необходимо только заменить  $A^T$  на  $A$ .

Что же касается условий падения или сохранения ранга возмущённой матрицы, когда вектор  $a$  зависит от столбцов, а вектор  $b^T$  - от строк  $A$ , то они имеют вид:

$b^T A^+ a \neq -1$  сохраняется,  $b^T A^+ a = -1$  падает.

24. *Квадрат расстояния вектора до гиперплоскости.* Квадрат расстояния  $\rho^2(a, \Gamma(b, L_A))$  вектора  $a \in R^m$  от гиперплоскости  $\Gamma(b, L_A) = b + L_A \subseteq R^m$  определяется соотношением

$$\rho^2(a, \Gamma(b, L_A)) = \min_{y \in \Gamma(b, L_A)} \|a - y\|^2 = (a - b)^T Z(A^T)(a - b).$$

Отметим, что привязка подпространств в этом, как и в других случаях, к множеству значений оператора  $A$ , не ограничивает сферу применимости результатов. С помощью п.1 они очевидным образом распространяются на ситуацию, когда подпространство порождается конечной совокупностью векторов.

### **Базовые нелинейные структуры: группирующие операторы**

Важнейшими нелинейными структурами евклидова пространства являются квадратичные формы (в работе - неотрицательно определённые) и отвечающие им эллипсы или эллипсоидальные цилиндры. Среди таких нелинейных структур принципиальными являются матрицы так называемых «группирующих операторов», которые естественным образом связаны с групповыми свойствами набора векторов. Группирующие операторы возникают в связи с набором векторов  $a(j) \in R^m, j = \overline{1, n}$ , и отвечающей ему матрицей  $A = (a(1) : \dots : a(n))$ . Как и ортогональные проекторы, группирующие операторы являются парными. Будем обозначать их, соответственно,  $R(A), R(A^T)$ .

25. *Определение группирующих операторов.* Группирующие операторы, обозначаемые  $R(A), R(A^T)$ , определяются соотношениями

$$R(A^T) = A^{+T} A^+, R(A) = A^+ A^{+T}.$$

Их важность для практики и область применения раскрываются в свойствах, приведённых ниже.

26. Проектирование на нормированный вектор  $u \in R^m : \|u\| = 1$ , элементов набора векторов  $a(j) \in R^m, j = \overline{1, n}$ . Основным результатом этого пункта представлен леммой 1 ниже.

Лемма 1. Для произвольного набора  $a(j) \in R^m, j = \overline{1, n}$ , с матричным представлением  $A = (a(1) : \dots : a(n))$  и произвольного нормированного вектора  $u \in R^m : \|u\| = 1$ , справедливо равенство:

$$\sum_{j=1}^n a^T(j) u u^T a(j) = u^T A A^T u \quad (14)$$

*Доказательство.* Действительно, принимая во внимание связь п.2 векторов набора  $a(j) \in R^m, j = \overline{1, n}$ , со своим матричным представлением:  $a(j) = A e_{(j)}, j = \overline{1, n}$ , имеем:

$$\sum_{j=1}^n a^T(j) u u^T a(j) = \sum_{j=1}^n e_{(j)}^T A^T u u^T A e_{(j)} = \sum_{j=1}^n u^T A e_{(j)} e_{(j)}^T A^T u = u^T A \left[ \sum_{j=1}^n e_{(j)} e_{(j)}^T \right] A^T u.$$

Остаётся только заметить, что  $\sum_{j=1}^n e_{(j)} e_{(j)}^T = E_n$ , где  $E_n$  - единичная матрица в  $R^n$ .

*Замечание 4.* Левая часть соотношения (14) леммы 1 представляет собою сумму квадратов проекций векторов набора  $a(j) \in R^m, j = \overline{1, n}$ , на нормированный вектор  $u \in R^m : \|u\| = 1$ .

27. Проектирование на элементы  $u_i \in R^m, i = \overline{1, r}$ , SVD – разложения матричного представления  $A$  набора векторов  $a(j) \in R^m, j = \overline{1, n}$ . Основным результатом этого пункта представлен леммой 2, приведённой ниже.

Лемма 2. Для произвольного набора  $a(j) \in R^m, j = \overline{1, n}$ , с матричным представлением  $A = (a(1) : \dots : a(n))$  имеет место соотношение:

$$\sum_{j=1}^n a^T(j) u_i u_i^T a(j) = \lambda_i^2, i = \overline{1, r}.$$

*Доказательство* вытекает из леммы 1 предыдущего пункта и из п.8, в котором наборы  $u_i \in R^m, v_i \in R^n, i = \overline{1, r}, r = \text{rank} A$ , определяются как ортонормированные наборы собственных векторов матриц  $A A^T, A^T A$ , отвечающих общему набору ненулевых собственных чисел  $\lambda_i^2 > 0, i = \overline{1, r}$ .

28. Группирующие операторы: эллипсы группировки набора векторов  $a(j) \in R^m, j = \overline{1, n}$ . Основное утверждение пункта – теорема 6 ниже.

Теорема 6. Пусть  $a(j) \in R^m, j = \overline{1, n}$ , произвольный набор векторов из  $R^m$  с матричным представлением  $A = (a(1) : \dots : a(n)), \text{rank} A = r \leq \min(m, n)$ . Тогда все векторы набора принадлежат внутренности эллипса, точнее: эллипсоидального цилиндра, определяемого уравнением

$$x^T R(A^T) x = r, x \in R^m,$$

где,  $R(A^T)$ , группирующий оператор:  $R(A^T) = A^{+T} A^+$ .

*Доказательство.* Рассмотрим квадраты проекций векторов набора  $a(j), j = \overline{1, n}$ , на каждый из векторов  $u_i, i = \overline{1, r}$ , SVD-представления (2) матрицы  $A$ . Принимая во внимание ортонормированность набора  $u_i, i = \overline{1, r}$ , и обозначая квадраты проекций через  $\|Pr_{u_i} a(j)\|^2, i = \overline{1, r}, j = \overline{1, n}$ , очевидным образом имеем

$$\|Pr_{u_i} a(j)\|^2 = a^T(j)u_i u_i^T a(j), j = \overline{1, n}, i = \overline{1, r}.$$

Суммирование по всем векторам набора  $a(j), j = \overline{1, n}$ , и применение леммы 2 даёт

$$\sum_{j=1}^n a^T(j)u_i u_i^T a(j) = u_i^T A A^T u_i = \lambda_i^2, i = \overline{1, r}.$$

Таким образом, после деления обеих частей последнего соотношения на, соответственно,  $\lambda_i^2, i = \overline{1, r}$ , имеем

$$\sum_{j=1}^n \frac{a^T(j)u_i u_i^T a(j)}{\lambda_i^2} = \frac{u_i^T A A^T u_i}{\lambda_i^2} = 1, i = \overline{1, r},$$

т.е.

$$\sum_{j=1}^n \frac{a^T(j)u_i u_i^T a(j)}{\lambda_i^2} = 1, i = \overline{1, r}.$$

Свернув (просуммировав) последнее равенство по  $i = \overline{1, r}$ , получаем

$$\sum_{i=1}^m \sum_{j=1}^n \frac{a^T(j)u_i u_i^T a(j)}{\lambda_i^2} = \sum_{i=1}^m \frac{u_i^T A A^T u_i}{\lambda_i^2} = r.$$

Поменяв порядок суммирования в двойной сумме, имеем

$$\sum_{j=1}^n \sum_{i=1}^m \frac{a^T(j)u_i u_i^T a(j)}{\lambda_i^2} = \sum_{j=1}^n a^T(j) \sum_{i=1}^m \frac{u_i u_i^T}{\lambda_i^2} a(j) = r.$$

Далее, приняв во внимание, что

$$\sum_{i=1}^r \frac{u_i u_i^T}{\lambda_i^2} = A^{+T} A^+ = R(A^T),$$

получаем окончательно

$$\sum_{j=1}^n \frac{a^T(j)u_i u_i^T a(j)}{\lambda_i^2} = \sum_{j=1}^n a^T(j) \sum_{i=1}^m \frac{u_i u_i^T}{\lambda_i^2} a(j) = \sum_{j=1}^n a^T(j) A^{+T} A^+ a(j) = \sum_{j=1}^n a^T(j) R(A^T) a(j) = r.$$

Таким образом,

$$\sum_{j=1}^n a^T(j) R(A^T) a(j) = r \quad (15)$$

Поскольку  $R(A^T)$  – симметричная, неотрицательно определённая матрица, то следствием соотношения (15) является одновременное выполнение неравенств

$$a^T(j) R(A^T) a(j) \leq r, j = \overline{1, n} \quad (16)$$

Та же симметричность и неотрицательная определённость позволяет сделать вывод, что уравнение

$$x^T R(A^T) x = r, x \in R^m \quad (17)$$

определяет эллипс, точнее: эллипсоидальный цилиндр, в  $R^m$  с длинами  $\frac{1}{\lambda_i \sqrt{r}}, i = \overline{1, r}$  нетривиальных полуосей. Напомним, что  $r = \text{rank}A \leq \min(m, n)$ .

Таким образом, выполнение неравенства (16) для всех векторов набора  $a(j) \in R^m, j = \overline{1, n}$ , означает их одновременную принадлежность внутренности эллипсоидального цилиндра с уравнением (17), и доказательство теоремы завершено.

*Замечание 5.* В действительности неравенство (16) может давать существенное закругление «радиуса» эллипса. Так, при векторах  $a(j) \in R^m, j = \overline{1, n}$ , близких к ортогональным, очевидным образом, константу в правой части (16) можно выбрать близкой к 1.

29. Усиление результата об эллипсах группировки.

**Теорема 7.** Все векторы набора  $a(j) \in R^m, j = \overline{1, n}$ , с матричным представлением  $A = (a(1): \dots : a(n))$  принадлежат внутренности эллипсоидального цилиндра

$$x^T R(A^T) x = r_{\max}^2, r_{\max}^2 \leq r = \text{rank}A, x \in R^m, r_{\max}^2 = \max_{j=\overline{1, n}} a^T(j) R(A^T) a(j),$$

который будем называть минимальным эллипсом группировки для рассматриваемого набора векторов.

### Группирующие операторы и расстояние Махаланобиса

Группирующие операторы с точностью до скалярного множителя неожиданным образом оказываются связанными с расстоянием Махаланобиса [Mahalanobis, 1936] (см. также, например, [McLachlan, Geoffry, 1992]), и проясняют его суть. Напомним, что расстояние Махаланобиса (в дальнейшем PaM) является одним из способов определения степени принадлежности  $\rho^2(x, N)$  неслучайного вектора  $x \in R^m$  многомерному  $N(a, B)$  нормальному распределению, и определяется соотношением

$$\rho^2(x, N) = (x - a)^T B^{-1} (x - a). \quad (18)$$

Напомним также, что  $a \in R^m$  является математическим ожиданием, а симметричная неотрицательно определённая матрица  $B: B \in R^{m \times m}, B^T = B, B \geq 0$ , является матрицей ковариаций распределения. Неслучайный вектор  $x \in R^m$  интерпретируется как возможный элемент выборки.

Как правило, при использовании PaM предполагается, что наблюдаемое значение может относиться к одному из нескольких  $N_k, k = \overline{1, K}$  нормальных распределений  $N(a_k, B_k), k = \overline{1, K}$ . В этом случае для каждого из возможных распределений определяется своё PaM. В этом случае рассматривается набор расстояний вида (18):

$$\rho^2(x, N_k) = (x - a_k)^T B_k^{-1} (x - a_k), \quad (19)$$

а отнесение вектора  $x \in R^m$  к одному из  $K$  возможных нормальных распределений осуществляется по минимуму PaM'ов из (19).

Отметим для сравнения, что средний квадрат расстояния от неслучайного вектора  $x \in R^m$  до  $N(a, B)$  распределённой случайной величины  $\xi$  определяется выражением:

$$M \|\xi - x\|^2 = \text{tr}B + \|x - a\|^2,$$

а квадратичная форма  $(x - a)^T B (x - a)$  описывает дисперсию случайной величины  $(x - a)^T \xi$ .

Таким образом, квадратичная форма (18) не является адекватной с точки зрения описания естественных квадратических характеристик нормально распределённого случайного вектора на основе среднего.

Единственным обоснованием использования расстояния Махаланобиса является присутствие квадратичной формы из (18) в плотности распределения  $f(x), x \in R^m$ , определяемой соотношением

$$f(x) = (2\pi)^{-\frac{m}{2}} |B|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-a)^T B^{-1}(x-a)\right\}, x \in R^m.$$

Заканчивая знакомство с РаМ, отметим, что при его практическом использовании вместо параметров  $a, B$  используются их оценки  $\hat{a}, \hat{B}$  на основе выборки  $a(1), \dots, a(n) \in R^m$ , определяемые соотношениями

$$\hat{a} = \bar{a} = \frac{1}{n} \sum_{i=1}^n a(i), \hat{B} = \frac{1}{n} \sum_{i=1}^n (a(i) - \bar{a})(a(i) - \bar{a})^T. \quad (20)$$

В этом случае РаМ приобретает статистический вид

$$\hat{\rho}^2(x, N) = (x - \hat{a})^T \hat{B}^{-1} (x - \hat{a}). \quad (21)$$

Теорема 8. Статистический вид РаМ может быть представлен в виде

$$\hat{\rho}^2(x, N) = n(x - \bar{a})^T R(\tilde{A}^T)(x - \bar{a}),$$

где  $\tilde{A} = (a(1) - \bar{a}; \dots; a(n) - \bar{a})$ .

*Доказательство.* Действительно, из (20) непосредственно следует, что  $\hat{a} = \bar{a}$ . Из тех же соотношений, с использованием результата п.3 вытекает, что

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n (a(i) - \bar{a})(a(i) - \bar{a})^T = \frac{1}{n} \tilde{A} \tilde{A}^T \quad (22)$$

Если матрица  $\hat{B}$  невырожденная и определяется соотношением (22), то  $\hat{B}^{-1} = n(\tilde{A} \tilde{A}^T)^{-1}$ . Из п.8 векторного SVD-разложения для матрицы  $\tilde{A}$  вытекает, что

$$\tilde{A} \tilde{A}^T = \sum_{i=1}^n \lambda_i^2 u_i^T u_i \quad (23)$$

Из (23) в свою очередь следует, что матрица  $(\tilde{A} \tilde{A}^T)^{-1}$ , если она существует, определяется соотношением

$$(\tilde{A} \tilde{A}^T)^{-1} = \sum_{i=1}^m \frac{u_i^T u_i}{\lambda_i^2} \quad (24)$$

Как нетрудно убедиться, правая часть в (24) совпадает с  $R(\tilde{A}^T)$ . Таким образом, окончательно имеем:

$$\hat{B}^{-1} = nR(\tilde{A}^T),$$

что и завершает доказательство теоремы.

Очевидна, таким образом, эвристическая основа расстояния Махаланобиса и его жёсткая теоретико-вероятностная привязка. Последнее означает, что вне теории вероятностей и математической статистики и вне связи с многомерным нормальным распределением говорить о расстоянии Махаланобиса не имеет смысла.

В то же время группирующие операторы предоставляют возможности выявления групповых свойств векторов на основе минимальных эллипсов группировки в любой ситуации.

---

### Применения: кластеризация

Примеры применения базовых линейных структур в линейной регрессии, в линейных системах управления с дискретным временем, а также для специального класса функциональных сетей, обобщающих искусственные нейронные сети, можно найти в работах [Кириченко, Донченко, 2005], [Донченко et al 2010].

Специально остановимся на применении базовых структур в задачах кластеризации. Отметим, что ПДО расширяет возможности кластеризации, позволяя эффективно погружать классифицируемые объекты в подходящие подпространства или гиперплоскости п.1 или минимальные эллипсы группировки. Теория ПДО даёт возможность связывать подпространство, порождённое набором векторов, как и минимальный эллипс группировки, с подходящей матрицей, строящейся конструктивно и в аналитическом виде. Если объект связывается с гиперплоскостью, то её смещение – это, как правило, среднее по векторам порождающей совокупности, а подпространство – это подпространство значений матрицы, построенной из центрированных средним векторов порождающей совокупности, как из столбцов. В случае группирования минимальным эллипсом группировки его центр – среднее по векторам совокупности, а матрица квадратичной формы – это, с точностью до транспонирования, группирующий оператор для матрицы, столбцовым представлением которой являются центрированные средним векторы исходной совокупности. Результаты п.25 обеспечивают возможность конструктивного вычисления расстояний от объектов (подпространств или гиперплоскостей), ассоциируемых с порождающей совокупностью в то время, как п. 29 – конструктивно строить расстояния соответствия на основе минимальных эллипсов группировки. Недостаток места не даёт возможности остановиться на результатах, обеспечивающих рекуррентность вычисления необходимых характеристик. С подробностями можно ознакомиться, например, в работах: [Кириченко, Донченко, 2007], [Кириченко, Донченко, 2008]).

Как уже было отмечено, в качестве расстояний соответствия в задачах кластеризации могут быть использованы как расстояния от гиперплоскостей, так и расстояния на основе минимальных эллипсов группировки. Последние дают возможность строить расстояния соответствия в связи с использованием базовой нелинейной структуры. В сущности, речь идёт о том, что результат теоремы 7 можно использовать для определения расстояния соответствия, точнее его квадрата, обозначим его  $\rho^2(x, Kl)$ , между вектором  $x \in R^m$  и кластером  $Kl$ , порожденным обучающей выборкой  $a(j) \in R^m, j = \overline{1, n}$ .

Обозначим через  $\bar{a}$  среднее обучающей выборки  $a(j) \in R^m, j = \overline{1, n}$ , а через  $\tilde{A}$  матричное представление для векторов  $\tilde{a}(j) \in R^m, j = \overline{1, n}$ , первоначального набора, центрированных средним:

$$\begin{aligned}\bar{a} &= \frac{1}{n} \sum_{j=1}^n a(j), \\ \tilde{a}(j) &= a(j) - \bar{a}, \\ \tilde{A} &= (\tilde{a}(1) : \dots : \tilde{a}(n)).\end{aligned}$$

Тогда квадрат расстояния, определяемый соотношением

$$\rho^2(x, Kl) = \frac{1}{r_{\max}^2} (x - \bar{a})^T R(\tilde{A}^T)(x - \bar{a}),$$

является расстоянием, определяемым минимальным эллипсом группировки

$$(x - \bar{a})^T R(\tilde{A}^T)(x - \bar{a}) = r_{\max}^2 \quad (25)$$

Уравнение (25) определяет минимальный эллипс группировки векторов  $\tilde{a}(j) \in R^m, j = \overline{1, n}$ ,  $\bar{a}$ .

При наличии набора кластеров  $Kl, l = \overline{1, L}$ , расстояния соответствия определяются, соотношениями

$$\rho^2(x, Kl_l) = \frac{1}{r_{l, \max}^2} (x - \bar{a}_l)^T R(\tilde{A}_l^T)(x - \bar{a}_l), l = \overline{1, L}, \quad (26)$$

с очевидной детализацией обозначений в связи с употреблением соответствующего индекса. Отнесение к тому или иному кластеру осуществляется по минимуму расстояний (26).

Примеры эффективного применения расстояний соответствия вида (26) можно найти, например, в работах [Кириченко, Донченко, 2008], [Кириченко, Донченко, 2007], [Донченко, Омардибирова 2005].

---

## Заклучение

---

В работе предложена и обоснована концепция базовых структур евклидового пространства, как линейных, так и нелинейных. Изложены конструктивные способы описания и взаимного перехода от одних типов структур к другим. В числе других рассмотрены конструктивные способы порождения, описания и использования базовых структур. Упомянутая конструктивность обеспечивается применением как классических результатов ПдО, так и новыми результатами в этой области. В частности, в работе приведена теорема о сингулярном разложении матрицы как линейного оператора между матричными пространствами, доказана теорема свёртки, обеспечивающая эквивалентность векторного и матричного сингулярных и ПдО-матриц на основе соответствующих сингулярных разложений. Рассмотрены применения полученных результатов для построения кластеризации с разными вариантами расстояний соответствия.

---

## Литература

---

- [Донченко et al, 2010] Донченко В.,Кривонос Ю., Омардибиров В. Базовые структуры евклидовых пространств: конструктивные методы описания и использования/ New Trends in Classification and Data Mining. – ITHEA, Sofia, Bulgaria. -2010.- ISBN 978-954-16-0042-9.- P. 155-170..
- [Донченко. 2009] Донченко В.С. Неопределённость и математические структуры в прикладных исследованиях/ Human aspects of Artificial Intelligence International Book Series Information science & Computing.– Number 12. – Supplement to International Journal “Information technologies and Knowledge”. –Volume 3.–2009. – P. 9-18.
- [Донченко, Омардибиров В.С. 2005] Донченко В.С., Омардибиров В.Н. Технология классификации электронных документов с использованием теории возмущения псевдообратных матриц// Proceedings of the XI-th International Conference “Knowledge-Dialogue-Solution”. – June 20-30, Varna, 2005. – Volume 1. – С.223-226.
- [Кириченко, Донченко, 2008] В.С Кириченко Н.Ф. Донченко В.С. Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация / Artificial Intelligence and Decision Making.– International book series “INFORMATION SCIENCE&COMPUTING”, Number 7.– Sofia 2008.– P. 25-36.
- [Moore, 1920] Moore E.H. On the reciprocal of the general algebraic matrix // Bulletin of the American Mathematical Society. – 26, 1920. – P.394 -395.
- [Penrose, 1955] Penrose R. A generalized inverse for matrices // Proceedings of the Cambridge Philosophical Society 51, 1955. – P.406-413.
- [Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание. – М.: Наука. – 1977.– 305 с.
- [Кириченко, 1997] Кириченко Н.Ф. Аналитическое представление псевдообратных матриц //Киб. и СА.- №2. –1997.– С.98-122.
- [Кириченко, Донченко,2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв'язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.
- [Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации// Киб. и СА.- №4, 2007– С.98-122.
- [Кириченко, Донченко, 2008] Кириченко Н.Ф. Донченко В.С. Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация / Artificial Intelligence and Decision Making.– International book series “INFORMATION SCIENCE&COMPUTING”, Number 7.– Sofia 2008. – P. 25-36.
- [Mahalanobis,1936] Mahalanobis, P C. On the generalized distance in statistics.//Proceedings of the National Institute of Sciences of India.- 1936.-2 (1).-P. 49–55.
- [McLachlan, Geoffry, 1992] McLachlan, Geoffry J. Discriminant Analysis and Statistical Pattern Recognition. Wiley Nescience.-1992.- ISBN 0471691151.

---

## Информация об авторе

---

**Владимир С. Донченко** – профессор; Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, e-mail: [voldon@unicyb.kiev.ua](mailto:voldon@unicyb.kiev.ua)



## МУЛЬТИМНОЖЕСТВА – АЛЬТЕРНАТИВА ТЕОРЕТИКО-МНОЖЕСТВЕННОЙ ПЛАТФОРМЫ В МАТЕМАТИЧЕСКИХ ОСНОВАНИЯХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Дмитрий Буй, Юлия Богатырёва

**Аннотация:** Основной тезис работы: мультимножества (совокупности с повторениями) могут рассматриваться в качестве альтернативы теоретико-множественной платформы в математических основаниях информационных технологий. В статье в систематизированном виде приводится обзор существующей литературы по теории и применению мультимножеств, рассматривается использование мультимножеств в ДНК-вычислениях. Введены основные определения, касающиеся теории мультимножеств. Построен содержательный фрагмент математической теории мультимножеств: частично упорядоченные семейства мультимножеств, решетки и полные решетки мультимножеств. Все это служит обоснованием рассмотрения мультимножеств в качестве упомянутой альтернативы.

**Ключевые слова:** математические основания информационных технологий, мультимножество, ДНК-вычисления, решетка мультимножеств, полная решетка мультимножеств.

**ACM Classification Keywords:** F.4.1 Mathematical Logic, G.2.1 Combinatorics, H.2.4 Systems – Relational databases.

---

### Введение

Возрастающие требования к информационным технологиям могут быть удовлетворены только созданием адекватных логико-математических основ; это, в свою очередь, приводит к пересмотру классической канторовской теоретико-множественной платформы современной математики. В статье в качестве альтернативы предлагается мультимножественная платформа.

Содержательно говоря, мультимножества – это совокупности с повторениями. Понятие мультимножества появилось относительно недавно, хотя в практических задачах мультимножества использовались довольно часто.

В 60-х годах известный специалист в области информатики Д.Кнут поставил вопрос об отсутствии адекватной терминологии и обозначений для такой глобальной концепции как мультимножество. Впервые термин “мультимножество” (multiset, bag) был предложен Н.Г. де Брейном (N. G. de Bruijn) в частной корреспонденции с Д. Кнутом. В 70-х годах этот термин широко распространился и сейчас является стандартным термином.

Мультимножества удобно использовать в качестве модели для представления и исследования объектов, особенностью которых является, во-первых, множественность и, во-вторых, повторяемость данных.

Широкое применение мультимножеств при решении теоретико-прикладных задач, в частности, задач информатики (табличные базы данных, теория принятия решений, теория информации и кодирования) вызывает необходимость расширения и уточнения соответствующих аспектов теории мультимножеств.

В статье приводится обзор всей доступной современной библиографии по мультимножествам (28 источников), строится фрагмент содержательной математической теории мультимножеств (исследование структуры семейства мультимножеств с естественным упорядочением, в частности, построение решетки мультимножеств и ее вложений в полные решетки).

---

## Библиография мультимножеств

---

Всю библиографию, посвященную мультимножествам, можно условно разделить на такие категории: работы, касающиеся абстрактной теории мультимножеств, обзорные работы, работы касающиеся использования мультимножеств. Кроме того отделим работы по применению теории мультимножеств именно в табличных базах данных. Следовательно, имеем 4 раздела, которые представлены ниже.

### I. АБСТРАКТНАЯ ТЕОРИЯ МУЛЬТИМНОЖЕСТВ

1. [Albert, 1991]. Автор дает формальное определение мультимножества и операциям над ними; приводит алгебраические свойства мультимножеств.
2. [Кнут, 2000]. Автор дает содержательное определения понятию мультимножества; определяет операции объединения, пересечения и суммы мультимножеств.
3. [Syropoulos, 2001]. Работа состоит из нескольких частей и подытоживает все то, что связано с теорией мультимножеств по состоянию на 2001 год. Сначала автор дает определение мультимножеству, вводит операции сложения, объединения и пересечения мультимножеств, а также понятия мощности и включения мультимножеств. Далее рассматриваются гибридные множества, нечеткие и частично упорядоченные мультимножества, а также приводится категориальная модель мультимножеств.
4. [Петровский, 2002], [Петровский, 2003] В первой монографии автор вводит основные определения теории мультимножеств: мультимножество, характеристическая функция, операции над мультимножествами. Кроме того, он рассматривает свойства основных операций над мультимножествами, методы графического представления мультимножеств и дает краткий обзор применения мультимножеств в различных областях. Во второй монографии автор рассматривает метрические пространства множеств и мультимножеств, описывает новые виды метрик.

### II. ОБЗОРНЫЕ СТАТЬИ

5. [Blizard, 1989]. Автор дает развернутый обзор теории мультимножеств по состоянию на 1989 год. Работа состоит из двух частей: теория мультимножеств и ее применение. В первой части работы, начиная от Кантора и его определения понятие множества, автор приводит разнообразные определения понятия мультимножества разных авторов. Во второй части работы мультимножества рассматриваются в первую очередь как объекты некоторых практических задач.
6. [Singh, 2007]. В работе рассматриваются разнообразные представления мультимножеств (в мультипликативной, линейной формах, в виде последовательности, как семейство множеств, в виде числовой последовательности). Определяются операции над мультимножествами и рассматриваются некоторые их свойства, а также дается краткий обзор применений мультимножеств в математике, компьютерные науках и других сферах.

### III. МУЛЬТИМНОЖЕСТВА В ТАБЛИЧНЫХ БАЗАХ ДАННЫХ

7. [Libkin, 1993], [Libkin, 1997]. В первой работе авторы рассматривают теоретические вопросы, касающиеся реляционных баз данных, основой которых выступают мультимножества. Они строят язык запросов BQL (Bag Query Language) и исследуют связь между полученным языком и так называемой вложенной реляционной алгеброй (nested relation algebra). Вторая работа посвящена выразительной силе языка запросов для мультимножеств, а также использованию некоторых конструкций для мультимножеств, множеств и списков.
8. [Буй, Поляков, 1999]. В работе задана композиционная семантика таблиц с дубликатами строк и таблиц, "упорядоченных" конструкцией ORDER BY.

9. [Редько, 2001]. Монография посвящена табличным алгебрам и SQL-подобным языкам. В ней дается формальное определение мультимножества, характеристической функции, а также определяются операции над мультимножествами: объединение, пересечение, разности (вычитания), декартового соединения, фильтрации, полного образа, агрегатные функции.

10. [Lamperti, 2001]. Статья посвящена расширению возможностей баз данных за счет использования мультимножеств. Авторы отмечают, что современные коммерческие реляционные базы данных предоставляют возможность осуществлять мультимножественно-ориентированные манипуляции над таблицами, даже если базы данных основаны на формальной множественно-ориентированной модели.

11. [Кузнецов]. Рассматривается существование такого структурного типа как мультимножество (BAG) в декларативном языке ограничений OCL. Это тип является разновидностью коллекций и имеет соответствующие операции.

12. [SQL:2003]. Начиная со стандарта SQL:2003 в язык SQL был введен конструктор типа MULTISSET. Значения мультимножеств задаются использованием специальной конструкции multiset value constructor. Кроме этого для мультимножеств введены операции объединения, пересечения и разности (multiset union, multiset intersect, multiset except), а также новые агрегатные функции (collect, fusion, intersect).

13. [Ross, 2004]. Авторы представляют симметрическую связь между  $k$ -арными сущностями базы данных как мультимножество мощности  $k$ , где  $k$  – натуральное число. В статье обосновывается необходимость поддержки базами данных мультимножеств, ограниченных по мощности (cardinality-bounded multisets), которые естественным образом возникают при решении реальных задач. Также предлагаются методы реализации. Описан синтаксис расширения SQL, что дает возможность формулировать запросы над такими симметричными связями.

14. [Гарсия-Молина, 2004]. Дается определение мультимножества в терминах табличных баз данных. Также над мультимножествами вводятся основные (объединение, пересечение, разность, проекция, выбор из мультимножества или селекция, декартовое сложение) и дополнительные (агрегирования, сортировки, группировки) операции.

#### IV. ДРУГИЕ ПРИМЕНЕНИЯ МУЛЬТИМНОЖЕСТВ

15. [Барендрегт, 1985]. При рассмотрении сильно эквивалентных редукций теории  $\lambda$ -исчислений вводятся мультимножества, основами которых выступают множества натуральных чисел, для таких мультимножеств вводятся ординалы.

16. [Knuth, 1992]. В своей работе автор определяет мультитязык как мультимножество строк и строит контекстно-свободный мультитязык.

17. [Lloyd, 1998-1], [Lloyd, 1998-2]. Автор предлагает новый способ поддержки мультимножеств в декларативном языке программирования Escher. Вводит стандартное определение мультимножества, определяет мультимножество стандартными средствами. Реализует операции сложения, объединения, пересечения и разности над мультимножествами, вспомогательные функции для работы с ними.

18. [Башкин, 2005], [Сети Петри, 2000]. Мультимножества используются для определения основных понятий сетей Петри.

19. [Сухольский, 2004]. Методы психологи используют мультимножества.

20. [Bonchis, 2005]. Статья посвящена представлению и кодированию информации в терминах теории мультимножеств. Информационный ресурс порождает мультимножественные сообщения. Исследуется норма энтропии мультимножественного информационного ресурса.

21. [Малинец, 2005]. Мультимножества используются в такой новой области знаний как вычисления на ДНК – разделе так называемых молекулярных вычислений (нового междисциплинарного направления на границе молекулярной биологии и компьютерных наук).

22. [Славин, 2006]. Мультимножества применяются в задачах распознавания символов. Разнообразные, с точки зрения изображения, типы символов могут содержать несколько графем, т.е. типов изображений, которые отвечают одному символу. Алфавит обучения как множество классов является носителем мультимножества всех допустимых графем. Кратность элемента этого множества есть количество графем, неразличимых с точки зрения алфавита обучения.

---

### Построение формальной модель ДНК-вычислений

---

Применение мультимножеств в ДНК-вычислениях рассмотрим более детально.

ДНК-вычисления – это форма вычислений, использующая ДНК, биохимию и молекулярную биологию вместо традиционных компьютерных технологий, основанных на кремнии [DNA]. Соответственно, ДНК-компьютер – вычислительная система, использующая вычислительные возможности молекул ДНК. Развитие этой предметной области началось в 1994 году работой Леонарда Адлемана (Leonard Adleman), в которой показано как с помощью пробирки, содержащей ДНК, можно весьма эффективно решить классическую так называемую задачу коммивояжера. Также известны работы, например, Эуда Шапиро (Ehud Shapiro) по реализации конечных автоматов, Эрика Винфри (Erik Winfree) по синтезу различных поверхностей (в частности, известной фрактальной структуры – ковра Серпинского) [Малинец, 2005].

Молекулу ДНК формально представим в виде пары слов в четырехбуквенном алфавите  $\aleph = \{A, T, C, G\}$ , каждый символ которого обозначает соответственно основания нуклеотидов – аденин, тимин, цитозин и гуанин:  $(\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n)$ , где  $\xi_1, \dots, \xi_n \in \aleph$ ,  $\eta_1, \dots, \eta_n \in \aleph$ . Если комплементарность уточнить в виде биекции  $\psi : \aleph \rightarrow \aleph$ ,  $\psi(A) = T$ ,  $\psi(T) = A$ ,  $\psi(C) = G$ ,  $\psi(G) = C$ , то тогда должно выполняться  $\psi(\xi_i) = \eta_i$  для всех  $i = 1, \dots, n$ . Таким образом, второе слово пары однозначно восстанавливается по первому слову, и вычисления на ДНК уточняются как вычисления над словами в указанном алфавите.

Формальная модель такой словарной вычислимости была построена в работе [Буй, Редько, 1984] в терминах так называемых примитивных программных алгебр (ППА).

Рассмотрим более сложный случай, когда, во-первых, цепочки ДНК имеют, вообще говоря, разную длину, и, во-вторых, учитывается “сдвиг” одной цепочки относительно другой (сдвиги возникают в виду удлинений, дополнений, укорочений, разрезов, сшивки цепочек): моделью ДНК является тройка вида  $(\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m, k)$ , где  $k \leq \min(n, m)$  – количество пар (нуклеотидов), связанных между собой. Формально говоря, должны выполняться утверждения: для всех  $i = 1, \dots, k$   $\xi_i = \psi(\eta_i(m - k + 1))$  или  $\eta_i = \psi(\xi_i(n - k + 1))$ . Первое равенство описывает случай сдвига первой (верхней) цепочки ДНК вправо, а второй – влево. В частном случае при  $n = m = k$  получается рассмотренный ранее случай цепочек равных длин без сдвигов. Решение вопроса вычислимой полноты для общего случая в терминах ППА требует отдельного рассмотрения.

Очевидно, что возникает потребность в общих моделях молекулярных вычислений, которые бы позволяли проектировать новые эксперименты и обобщать существующие. Одна из таких моделей – модель параллельной фильтрации (Parallel Filtering Model). Основой такой модели выступает “пробирка”, формальной моделью которой, в свою очередь, является мультимножество слов.

Класс вычислимых функций над мультимножествами задается в виде замыкания семейства простейших функций и предикатов над мультимножествами в ППА. Такое семейство состоит из константной функции, фиксирующей пустое мультимножество, функций сложения и вычитания мультимножеств, функции выбора компоненты мультимножества, предиката равенства мультимножеств и некоторых иных специфических функций над мультимножествами.

### Основные определения

Введем формальное определение мультимножества. Мультимножество  $\alpha$  с основой  $U$  – это функция вида  $\alpha : U \rightarrow N^+$ , где  $U$  – некоторое множество (в классическом канторовском понимании), а  $N^+ = \{1, 2, \dots\}$  – множество натуральных чисел без нуля [Петровский, 2002], [Редько, 2001]. Пусть задано мультимножество  $\alpha$  с основой  $U_\alpha = \text{dom } \alpha$ . Здесь  $\text{dom } \alpha$  – область определения мультимножества как функции. Характеристической функцией мультимножества  $\alpha$  называется функция вида  $\chi_\alpha : D \rightarrow N$ , значение которой задается кусочной схемой

$$\chi_\alpha(d) = \begin{cases} \alpha(d), & \text{если } d \in \text{dom } \alpha, \\ 0, & \text{иначе;} \end{cases}$$

для всех  $d \in D$ , где  $D$  – универсум элементов основ мультимножеств [Петровский, 2002], [Редько, 2001]. Очевидно, что по характеристической функции соответствующее мультимножество восстанавливается однозначно и операции над мультимножествами удобно задавать характеристическими функциями.

Введем бинарное отношение включения на мультимножествах: мультимножество  $\beta$  включается в мультимножество  $\alpha$  ( $\beta \leq \alpha$ ), если для их характеристических функций выполняется утверждение  $\chi_\beta(d) \leq \chi_\alpha(d)$ ,  $\forall d \in D$ . Непосредственно проверяется, что отношение включения является частичным порядком.

Пусть  $M$  – семейство мультимножеств соответствующего универсума  $D$ . Рассмотрим частично упорядоченное множество (ч.у.м.)  $\langle M, \leq \rangle$  с введенным выше порядком включения и ч.у.м. характеристических функций  $\langle \{\chi_\alpha \mid \alpha \in M\}, \leq \rangle$ . Порядок  $\leq$  определяется следующим образом:

$\chi_\alpha \leq \chi_\beta \stackrel{\text{def}}{\Leftrightarrow} \forall d (d \in D \Rightarrow \chi_\alpha(d) \leq \chi_\beta(d))$ . Второе ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \leq \rangle$  является прямым произведением множества натуральных чисел  $N$  со стандартным порядком  $\leq$ , т.е.  $\{\chi_\alpha \mid \alpha \in M\} = \prod_{d \in D} N$ . Нетрудно показать, что эти ч.у.м. изоморфны. Следовательно, структура ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \leq \rangle$  сохраняется на ч.у.м.  $\langle M, \leq \rangle$ . А свойства ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \leq \rangle$  вытекают из свойств ч.у.м.  $\langle N, \leq \rangle$ .

Введем операции объединения и пересечения мультимножеств [Syropoulos, 2001], [Петровский, 2002], [Редько, 2001]. Операция  $\cup_{\text{All}}$  мультимножеств  $\alpha$  и  $\beta$  сопоставляет мультимножество  $\alpha \cup_{\text{All}} \beta$ , значение характеристической функции которого на произвольном аргументе  $d$  задается выражением  $\max(\chi_\alpha(d), \chi_\beta(d))$ . Операция  $\cap_{\text{All}}$  мультимножеств  $\alpha$  и  $\beta$  сопоставляет мультимножество  $\alpha \cap_{\text{All}} \beta$ , значение характеристической функции которого на произвольном аргументе  $d$  задается выражением  $\min(\chi_\alpha(d), \chi_\beta(d))$ .

### Построение решетки мультимножеств

Рассмотренные выше операции объединения и пересечения мультимножеств имеют стандартные свойства.

**Лемма 1.** Операции  $\cup_{All}$  и  $\cap_{All}$  идемпотентны (т.е.  $\alpha \cup_{All} \alpha = \alpha$ ,  $\alpha \cap_{All} \alpha = \alpha$ ), коммутативны и ассоциативны.  $\square$

Доказательство вытекает из того, что теоретико-числовые операции  $\max$ ,  $\min$  имеют те же самые свойства.  $\square$

Таким образом, можно рассматривать две коммутативные идемпотентные полугруппы  $\langle M, \cup_{All} \rangle$  и  $\langle M, \cap_{All} \rangle$ , где  $M$  – семейство мультимножеств соответствующего универсума  $D$ .

Используя результат теории решеток (см. например, [Скорняков, 1986], § 8, с. 151, теорема 1), можно полугруппу по объединению превратить в верхнюю полурешетку, а полугруппу по пересечению – в нижнюю. Частичные порядки верхней и нижней полурешеток задаются соответственно следующими определениями:  $\alpha \bar{\leq} \beta \stackrel{def}{\Leftrightarrow} \alpha \cup_{All} \beta = \beta$ ,  $\alpha \bar{\leq} \beta \stackrel{def}{\Leftrightarrow} \alpha \cap_{All} \beta = \alpha$ , причем  $\sup_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cup_{All} \beta$ ,  $\inf_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cap_{All} \beta$ . Непосредственно проверяется, что эти порядки совпадают с введенным ранее порядком включения мультимножеств  $\preceq$ .

**Теорема 1.** Частично упорядоченное множество  $\langle M, \bar{\leq} \rangle$  является решеткой, причем  $\sup_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cup_{All} \beta$ ,  $\inf_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cap_{All} \beta$ .  $\square$

Такой способ построения решетки мультимножеств явно не использовал законы поглощения. Покажем, в общем случае, их роль при построении решетки по двум коммутативным идемпотентным полугруппам, сигнатурные операции которых связаны законами поглощения. Рассмотрим две коммутативные идемпотентные полугруппы  $\langle A, + \rangle$  и  $\langle A, \cdot \rangle$ , где  $A$  – некоторое абстрактное множество.

Используя хорошо известный результат теории решеток о связи коммутативных идемпотентных полугрупп и полурешеток (полуструктур) [Скорняков, 1986, § 8, с. 151, теорема 1], полугруппу  $\langle A, + \rangle$  можно превратить в верхнюю полурешетку, а полугруппу  $\langle A, \cdot \rangle$  – в нижнюю. Соответствующие частичные порядки верхней и нижней полурешеток задаются выражениями:

$$a \leq b \Leftrightarrow a + b = b, \quad a < b \Leftrightarrow ab = a,$$

причем  $\sup_{\leq} \{a, b\} = a + b$ ,  $\inf_{<} \{a, b\} = ab$ . Заметим, что, согласно стандартным соглашениям [Мальцев, 1970], [Скорняков, 1986], знак “ $\cdot$ ” в выражениях опускается.

**Теорема 2** (критерий совпадения порядков верхней и нижней полурешеток). Частичные порядки верхней и нижней полурешеток совпадают тогда и только тогда, когда выполняются законы поглощения.  $\square$

Доказательство. Сначала докажем необходимость. Если порядки совпадают, то заданное множество является одновременно и верхней и нижней полурешеткой, а, значит, является решеткой. Для решеток же законы поглощения выполняются [Мальцев, 1970], [Скорняков, 1986].

Докажем достаточность. Для этого нужно показать, что для  $\forall a, b$  выполняется эквивалентность:  $a + b = b \Leftrightarrow ab = a$ .

Допустим, что равенство  $a + b = b$  выполняется. Тогда  $ab = a(a + b)$ . По закону поглощения  $a(a + b) = a$ , поэтому  $ab = a$ .

Аналогично сделаем допущение, что  $ab = a$ . В этом случае  $a + b = ab + b$ . По закону поглощения  $ab + b = b$ , значит,  $a + b = b$ . □

Естественно, этот общий результат применим к построению решетки мультимножеств. Для этого надо убедиться в выполнении законов поглощения:  $\alpha \cap_{All} (\alpha \cup_{All} \beta) = \alpha$  и  $\alpha \cup_{All} (\alpha \cap_{All} \beta) = \alpha$ , что делается непосредственно (и, в свою очередь, вытекает из выполнения законов поглощения для теоретико-числовых функций  $\max$ ,  $\min$ ).

### Построение полной решетки мультимножеств

Приведем более сильные результаты о структуре введенного частично упорядоченного семейства мультимножеств.

**Предложение 1** (структура ч.у.м. мультимножеств). Выполняются следующие утверждения:

1. пустое мультимножество  $\emptyset_m$  (его характеристической функцией является константная функция, всюду равная нулю) – наименьший элемент в  $\langle M, \preceq \rangle$ ; 2.  $\inf \mu = \alpha$ , для произвольного непустого множества мультимножеств  $\mu \subseteq M$ ; здесь характеристическая функция мультимножества  $\alpha$  задается выражением  $\chi_\alpha(d) = \min_{\beta \in \mu} \chi_\beta(d)$  (отметим, что инфимум пустого множества мультимножеств не существует, так как в рассматриваемом частично упорядоченном множестве мультимножеств не существует наибольшего элемента); 3. для произвольного множества мультимножеств  $\mu$ : супремум  $\mu$  существует  $\Leftrightarrow \mu$  ограничено сверху; 4.  $\sup \mu = \alpha$ , где  $\mu$  – произвольное множество мультимножеств, имеющее точную верхнюю грань, а характеристическая функция мультимножества  $\alpha$  задается выражением  $\chi_\alpha(d) = \max_{\beta \in \mu} \chi_\beta(d)$ . □

Доказательство. Рассмотрим отдельно каждое из сформулированных выше утверждений.

1. Наименьшим элементом ч.у.м.  $\langle N, \leq \rangle$  является число 0. Соответственно, наименьшим элементом ч.у.м. характеристических функций  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  будет функция, всюду равная нулю. Исходя из того, что ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  изоморфно ч.у.м.  $\langle M, \preceq \rangle$ , получим, что наименьшим элементом второго будет пустое мультимножество  $\emptyset_m$ . 2. Доказательство второго утверждения вытекает из известной формулы нахождения инфимума множества функций (см., например, [Буй, 2002] или [Бурбаки, 1965]) и того факта, что ч.у.м.  $\langle N, \leq \rangle$  является вполне упорядоченным. 3. Супремум подмножества  $L$  ч.у.м.  $\langle N, \leq \rangle$  существует тогда и только тогда, когда множество  $L$  – конечно (ограничено). Перенесем этот результат на ч.у.м. характеристических функций, а, следовательно, и на ч.у.м. мультимножеств (так как они изоморфны). Тогда получим, что супремум произвольного подмножества  $\mu$  ч.у.м.  $\langle M, \preceq \rangle$  существует, тогда и только тогда, когда это множество ограничено сверху. 4. Доказательство этого утверждения также вытекает из известной формулы нахождения супремума множества функций при условии, что супремум существует. □

Таким образом, согласно предложению 1, ч.у.м.  $\langle M, \preceq \rangle$  имеет наименьший элемент (пустое мультимножество –  $\emptyset_m$ ) и любое его подмножество, ограниченное сверху, имеет точную верхнюю грань. Значит, следуя терминологии работ [Davey, 1990], [Биркгоф, 1984], ч.у.м.  $\langle M, \preceq \rangle$  является одновременно

условно полным множеством и полной полурешеткой (complete semilattice). Таким образом, установлена

**Теорема 3.** Ч.у.м.  $\langle M, \preceq \rangle$  является условно полным множеством и полной полурешеткой, при этом точные грани находятся согласно формул предложения 1.  $\square$

Доказательство вытекает из определений условно полного множества и полурешетки и предложения 1.  $\square$

Пополним частично упорядоченное множество  $\langle M, \preceq \rangle$  наибольшим элементом  $T$ ; полученное ч.у.м. обозначим через  $\langle M \cup \{T\}, \preceq \rangle$ .

**Следствие 1.** Ч.у.м.  $\langle M \cup \{T\}, \preceq \rangle$  является полной решеткой с наименьшим элементом  $\emptyset_m$  и наибольшим элементом  $T$ .  $\square$

Доказательство проводится очевидным образом; кроме того, можно использовать общий результат теории решеток [Биркгоф, 1984].  $\square$

Ч.у.м.  $\langle M, \preceq \rangle$  можно вложить и в другую полную решетку. Для этого расширим понятие мультимножества.

С этой целью пополним множество натуральных чисел без нуля  $N^+$  наибольшим элементом  $\infty$  и положим  $N_\infty^+ = N^+ \cup \{\infty\}$ . Под мультимножеством будем понимать функцию вида  $\alpha : U_\alpha \rightarrow N_\infty^+$ , семейство всех таких мультимножеств обозначим через  $M_\infty$ . Порядок на множестве  $N_\infty^+$  обозначим через  $\leq_\infty$ , тогда порядок на мультимножествах  $\preceq$  расширяется так:  $\alpha \preceq_\infty \beta \stackrel{def}{\Leftrightarrow} \forall d (\chi_\alpha(d) \leq_\infty \chi_\beta(d))$ ,  $\alpha, \beta \in M_\infty$ .

Отметим характерный факт: именно мультимножества с не более чем счетной кратности элементов основ естественно появляются при задании денотационной семантики рекурсивных запросов в SQL-подобных языках [Буй, Поляков, 2010].

**Теорема 4.** Ч.у.м.  $\langle M_\infty, \preceq_\infty \rangle$  является полной решеткой с наименьшим элементом  $\emptyset_m$  и наибольшим элементом  $T_\infty$ , где  $T_\infty : D \rightarrow \{\infty\}$ ,  $T_\infty(d) = \infty$  для всех  $d \in D$ . Точные нижние грани находятся по формулам предложения 1, для точных верхних граней выполняется формула  $\sup \mu = \alpha$ , где характеристическая функция мультимножества  $\alpha$  такая -  $\chi_\alpha(d) = \sup_{\mu \preceq_\infty \alpha} \chi_\mu(d)$ .  $\square$

Доказательство. Очевидно, что  $\langle N_\infty^+, \leq_\infty \rangle$  является полной решеткой. Остается применить тот хорошо известный факт, что прямое произведение полных решеток будет полной решеткой, и использовать уже неоднократно упоминавшуюся формулу для нахождения точных граней подмножеств прямого произведения.  $\square$

---

## Выводы

---

В статье в систематизированном виде представлен обзор существующей литературы по теории и применению мультимножеств (уточняющих понятие совокупности с конечными повторениями). Рассмотренные работы разделены на четыре раздела: абстрактная теория мультимножеств, обзорные статьи по мультимножествам, применение мультимножеств в табличных базах данных и другие применения мультимножеств.

В работе также приведены базовые определения, касающиеся теории мультимножеств: мультимножества, его характеристической функции, отношения включения, операций объединения и



пересечения. Построена решетка мультимножеств, выяснена структура ч.у.м. мультимножеств по естественному отношению включения. Построены вложение ч.у.м. мультимножеств в две полные решетки, при этом одна из этих полных решеток строится путем обобщения понятия мультимножества (допускается бесконечная кратность элементов основы).

Таким образом, мультимножества имеют многочисленные применения и содержательную математическую теории; следовательно, мультимножества действительно можно рассматривать в качестве альтернативы (по крайней мере потенциально) канторовской теоретико-множественной платформы и, значит, теорию мультимножеств необходимо вводить в учебные программы дисциплин, поддерживающих математические основы информатики (в частности, дискретная математика, теория табличных баз данных, биоинформатика, теория принятия решений).

---

### Список литературы

- [Albert, 1991] J. Albert. Algebraic properties of bag data types // Seventeenth International Conference on Very Large Data Bases. – 1991. – P. 211-219.
- [Blizard, 1989] W. Blizard. The Development of Multiset Theory // Notre Dame Journal of Formal Logic. – 1989. – Vol. 30, No. 1. – P. 36-66.
- [Bonchis, 2005] C. Bonchis, C. Izbasa, G. Ciobanu. Information Theory over Multiset // Research Institute “re-Austria”, Institute of Computer Science. – 2005.
- [Davey, 1990] B.A. Davey, H.A. Priestly. Introduction to Lattice and Order. – Cambridge: Cambridge University Press, 1990. – 248 p.
- [DNA] DNA computing. [Электронный ресурс]: from Wikipedia, the free encyclopedia. – Режим доступа: [http://en.wikipedia.org/wiki/DNA\\_computing](http://en.wikipedia.org/wiki/DNA_computing).
- [Knuth, 1992] D.E. Knuth. Context-Free Multilanguages // Theoretical Studies in Computer Science. – Academic Press. – 1992. – P. 1-13.
- [Lamperti, 2001] G. Lamperti, M. Melchiori, M. Zanella. On Multisets in Database Systems // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 147-215.
- [Libkin, 1997] L. Libkin, L. Wong. Query Language for Bags and Aggregates Function // Journal of Computer and System Sciences. – 1997. – Vol. 55, No. 1. – P. 241-272.
- [Libkin, 1993] L. Libkin, L. Wong. Some Properties of Query Language for Bags // Proceedings of 4th International Workshop on Database Programming Languages. – New York. – 1993. – P. 97-114.
- [Lloyd, 1998-1] J.W. Lloyd. Programming with Multisets // Department of Computer Science University of Bristol. – 1998.
- [Lloyd, 1998-2] J.W. Lloyd. Programming with Sets and Multisets // Department of Computer Science, Bristol. – 1998.
- [Ross, 2004] K.A. Ross, J. Stoyanovich. Symmetric relations and cardinality-bounded multisets in database systems // Very Large Database Endowment: international conference. – Toronto, Canada: proceedings. – 2004. – Vol. 30. – P. 912-923.
- [Singh, 2007] D. Singh, A.M. Ibrahim, T. Yohanna, J.N. Singh. An Overview of the Applications of Multisets // Novi Sad Journal of Mathematics. – 2007. – Vol. 37, No. 2. – P. 73-92.
- [SQL:2003] Наиболее интересные новшества в стандарте SQL:2003 [Электронный ресурс] – Режим доступа: <http://www.nestor.minsk.by/sr/2004/03/40331.html>.
- [Syropoulos, 2001] A. Syropoulos. Mathematics of Multisets // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 347-358.
- [Барендрегт, 1985] X. Барендрегт. Лямбда-исчисление.: [пер. с англ.]. – Москва: Мир, 1985. – 606 с.
- [Биркгоф, 1984] Г. Биркгоф. Теория решеток. – Москва: Наука, 1984. – 564 с.
- [Башкин, 2005] В.А. Башкин, И.А. Ломазова. Подобие обобщенных ресурсов в сетях Петри [Электронный ресурс] – Режим доступа: <http://lvk.cs.msu.su/files/mco2005/bashkin.pdf>.

- [Буй, 2002] Д.Б. Буй. Теорія програмних алгебр композиційного типу та її застосування: дисертація доктора фізико-математичних наук: 01.05.03 – математичне та програмне забезпечення обчислювальних машин та систем. – Київ, 2002. – 365 с.
- [Буй, Поляков, 1999] Д.Б. Буй, С.А. Поляков. Композиційна семантика SQL-подібних мов: мультимножини, рядки, впорядковані таблиці // Вісник Київського університету. Сер.: фіз.-мат. науки. – 1999. Вип. 2. – С. 183-194.
- [Буй, Поляков, 2010] Д.Б. Буй, С.А. Поляков. Композиційна семантика рекурсивних запитів в SQL-подібних мовах // Вісник Київського університету. Сер.: фіз.-мат. науки. – 2010. Вип. 1. – С. 45-50.
- [Буй, Редько, 1984] Д.Б. Буй, В.Н. Редько. Примитивные программные алгебры целочисленных и словарных функций // Докл. АН УССР, Сер. А. Физ.-мат. и техн. науки. – 1984. – № 10. – С. 69-71.
- [Бурбаки, 1965] Н. Бурбаки. Теория множеств. – Москва: Мир, 1965. – 455 с.
- [Гарсиа-Молина, 2004] Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. Системы баз данных: [полный курс: пер. с англ.] – Москва: Вильямс, 2004. – 1088 с.
- [Кнут, 2000] Д. Кнут. Искусство программирования: [2 том, 3-е изд.: пер. с англ.] – Москва: Вильямс, 2000. – 832 с.
- [Кузнецов] С.Д. Кузнецов. Концептуальное проектирование реляционных баз данных с использованием языка UML [Электронный ресурс] – Режим доступа: <ftp://ftp.dol.ru/pub/users/cgntv/download/sbornic/sbornic9/Doc13.doc>.
- [Малинец, 2005] Г.Г. Малинецкий, С.А. Науменко. Вычисления на ДНК. Эксперименты. Модели. Алгоритмы. Инструментальные средства [Электронный ресурс] – Режим доступа: [http://www.keldysh.ru/papers/2005/prep57/rpr2005\\_57.html](http://www.keldysh.ru/papers/2005/prep57/rpr2005_57.html).
- [Мальцев, 1970] А.И. Мальцев. Алгебраические системы. – Москва: Наука, 1970. – 392 с.
- [Петровский, 2002] А.Б. Петровский. Основные понятия теории мультимножеств. – Москва: Едиториал УРСС, 2002. – 80 с.
- [Петровский, 2003] А.Б. Петровский. Пространства множеств и мультимножеств. – Москва: Едиториал УРСС, 2003. – 248 с.
- [Редько, 2001] В.Н. Редько, Ю.И. Брона, Д.Б. Буй, С.А. Поляков. Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: Видавничий дім "Академперіодика", 2001. – 198 с.
- [Сети Петри, 2000] Сети Петри [Электронный ресурс] – Режим доступа: [http://www.iacp.dvo.ru/lab\\_11/otchet/ot2000/pn3.html#top](http://www.iacp.dvo.ru/lab_11/otchet/ot2000/pn3.html#top).
- [Славин, 2006] О.А. Славин. Использование мультимножеств в распознавании символов // Труды института системного анализа российской академии наук. – 2006. – Том. 23. – С. 198-205.
- [Скорняков, 1986] Л.А. Скорняков. Элементы алгебры. – Москва: Наука, 1986. – 240 с.
- [Сухольский, 2004] Г.В. Сухольский. Математические методы в психологии. – Харьков: ФОПІО, 2004. – 282 с.

---

### Информация об авторах

---

**Буй Дмитрий Борисович** – заведующий сектором проблем программирования, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: 01601, ул. Владимирская, 64, Киев, Украина; e-mail: [buy@unicyb.kiev.ua](mailto:buy@unicyb.kiev.ua)

**Богатырёва Юлия Александровна** – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: 01601, ул. Владимирская, 64, Киев, Украина; e-mail: [j\\_bogatyreva@ukr.net](mailto:j_bogatyreva@ukr.net)

## СТРУКТУРА ЧАСТИЧНО УПОРЯДОЧЕННОГО СЕМЕЙСТВА МУЛЬТИМНОЖЕСТВ

Дмитрий Буй, Юлия Богатырёва

**Аннотация:** Статья посвящена некоторым вопросам теории мультимножеств. Введены основные определения, касающиеся теории мультимножеств. Построена решетка мультимножеств. Показана роль законов поглощения при построении решетки мультимножеств. Кроме того, частичное упорядоченное множество мультимножеств вложено в две полные решетки, одна из которых построена путем обобщения понятия мультимножества (допускается бесконечная кратность элементов основы).

**Ключевые слова:** мультимножество, решетка мультимножеств, полная решетка мультимножеств.

**ACM Classification Keywords:** F.4.1 Mathematical Logic, G.2.1 Combinatorics, H.2.4 Systems – Relational databases.

### Введение

Мультимножества – это совокупности с повторениями. Мультимножества удобно использовать в качестве модели для представления и исследования объектов, особенностью которых является, во-первых, множественность и, во-вторых, повторяемость данных.

Анализ современной литературы по мультимножествам [Blizard, 1989], [Lamperti, 2001], [Singh, 2007], [Буй, Богатырёва 2010] свидетельствует о достаточно широком применении мультимножеств при решении теоретико-прикладных задач. Это вызывает необходимость расширения и уточнения соответствующих аспектов теории мультимножеств.

В статье рассматривается вопрос построения (полной) решетки мультимножеств.

### Основные определения

Введем формальное определение мультимножества. Мультимножество  $\alpha$  с основой  $U$  – это функция вида  $\alpha : U \rightarrow N^+$ , где  $U$  – некоторое множество (в классическом канторовском понимании), а  $N^+ = \{1, 2, \dots\}$  – множество натуральных чисел без нуля [Петровский, 2002], [Редько, 2001]. Зафиксируем некоторое множество  $D$ . Всюду далее будем рассматривать мультимножества, элементы основ которых принадлежат множеству  $D$ .

Пусть задано мультимножество  $\alpha$  с основой  $U_\alpha = \text{dom } \alpha$ . Здесь  $\text{dom } \alpha$  – область определения мультимножества как функции. Характеристической функцией мультимножества  $\alpha$  называется функция вида  $\chi_\alpha : D \rightarrow N$ , значение которой задается кусочной схемой

$$\chi_\alpha(d) = \begin{cases} \alpha(d), & \text{если } d \in \text{dom } \alpha, \\ 0, & \text{иначе;} \end{cases}$$

для всех  $d \in D$ , где  $D$  – как было определено выше, универсум элементов основ мультимножеств [Петровский, 2002], [Редько, 2001]. Очевидно, что по характеристической функции соответствующее мультимножество восстанавливается однозначно. Таким образом, операции над мультимножествами удобно задавать в терминах характеристических функций.

Введем бинарное отношение включения на мультимножествах: мультимножество  $\beta$  включается в мультимножество  $\alpha$  ( $\beta \leq \alpha$ ), если для их характеристических функций выполняется утверждение  $\chi_\beta(d) \leq \chi_\alpha(d)$ ,  $\forall d \in D$ . Непосредственно проверяется, что отношение включения является частичным порядком.

Пусть  $M$  – семейство мультимножеств соответствующего универсума  $D$ . Рассмотрим частично упорядоченное множество (ч.у.м.)  $\langle M, \leq \rangle$  с введенным выше порядком включения и ч.у.м. характеристических функций  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$ . Порядок  $\triangleleft$  определяется следующим образом:

$\chi_\alpha \triangleleft \chi_\beta \stackrel{\text{def}}{\Leftrightarrow} \forall d (d \in D \Rightarrow \chi_\alpha(d) \leq \chi_\beta(d))$ . Второе ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  является прямым произведением множества натуральных чисел  $N$  со стандартным порядком  $\leq$ , т.е.  $\{\chi_\alpha \mid \alpha \in M\} = \prod_{d \in D} N$ . Нетрудно показать, что эти ч.у.м. изоморфны. Следовательно, структура ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  сохраняется на ч.у.м.  $\langle M, \leq \rangle$ . А свойства ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  вытекают из свойств ч.у.м.  $\langle N, \leq \rangle$ .

Введем операции объединения и пересечения мультимножеств [Syropoulos, 2001], [Петровский, 2002], [Редько, 2001]. Операция  $\cup_{All}$  мультимножеств  $\alpha$  и  $\beta$  сопоставляет мультимножество  $\alpha \cup_{All} \beta$ , значение характеристической функции которого на произвольном аргументе  $d$  задается выражением  $\max(\chi_\alpha(d), \chi_\beta(d))$ . Операция  $\cap_{All}$  мультимножеств  $\alpha$  и  $\beta$  сопоставляет мультимножество  $\alpha \cap_{All} \beta$ , значение характеристической функции которого на произвольном аргументе  $d$  задается выражением  $\min(\chi_\alpha(d), \chi_\beta(d))$ .

### Построение решетки мультимножеств

Рассмотренные выше операции объединения и пересечения мультимножеств имеют стандартные свойства.

**Лемма 1.** Операции  $\cup_{All}$  и  $\cap_{All}$  идемпотентны (т.е.  $\alpha \cup_{All} \alpha = \alpha$ ,  $\alpha \cap_{All} \alpha = \alpha$ ), коммутативны и ассоциативны.  $\square$

Доказательство вытекает из того, что теоретико-числовые операции  $\max$ ,  $\min$  имеют те же самые свойства.  $\square$

Таким образом, можно рассматривать две коммутативные идемпотентные полугруппы  $\langle M, \cup_{All} \rangle$  и  $\langle M, \cap_{All} \rangle$ , где  $M$  – семейство мультимножеств соответствующего универсума  $D$ .

Используя результат теории решеток (см. например, [Скорняков, 1986], § 8, с. 151, теорема 1), можно полугруппу по объединению превратить в верхнюю полурешетку, а полугруппу по пересечению – в нижнюю. Частичные порядки верхней и нижней полурешеток задаются соответственно следующими определениями:

$$\alpha \bar{\leq} \beta \stackrel{\text{def}}{\Leftrightarrow} \alpha \cup_{All} \beta = \beta, \quad \alpha \underline{\leq} \beta \stackrel{\text{def}}{\Leftrightarrow} \alpha \cap_{All} \beta = \alpha,$$

причем  $\sup_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cup_{All} \beta$ ,  $\inf_{\underline{\leq}} \{\alpha, \beta\} = \alpha \cap_{All} \beta$ . Непосредственно проверяется, что эти порядки совпадают с введенным ранее порядком включения мультимножеств  $\leq$ .

**Теорема 1.** Частично упорядоченное множество  $\langle M, \leq \rangle$  является решеткой, причем  $\sup_{\underline{\leq}} \{\alpha, \beta\} = \alpha \cup_{All} \beta$ ,  $\inf_{\bar{\leq}} \{\alpha, \beta\} = \alpha \cap_{All} \beta$ .  $\square$

Такой способ построения решетки мультимножеств явно не использовал законы поглощения. Покажем, в общем случае, их роль при построении решетки по двум коммутативным идемпотентным полугруппам, сигнатурные операции которых связаны законами поглощения.

Рассмотрим две коммутативные идемпотентные полугруппы  $\langle A, + \rangle$  и  $\langle A, \cdot \rangle$ , где  $A$  – некоторое абстрактное множество.

Используя хорошо известный результат теории решеток о связи коммутативных идемпотентных

полугрупп и полурешеток (полуструктур) [Скорняков, 1986, § 8, с. 151, теорема 1], полугруппу  $\langle A, + \rangle$  можно превратить в верхнюю полурешетку, а полугруппу  $\langle A, \cdot \rangle$  – в нижнюю. Соответствующие частичные порядки верхней и нижней полурешеток задаются выражениями:

$$a \leq b \Leftrightarrow a + b = b, \quad a < b \Leftrightarrow ab = a,$$

причем  $\sup_{\leq} \{a, b\} = a + b$ ,  $\inf_{<} \{a, b\} = ab$ . Заметим, что, согласно стандартным соглашениям [Мальцев, 1970], [Скорняков, 1986], знак “ $\cdot$ ” в выражениях опускается.

**Теорема 2** (критерий совпадения порядков верхней и нижней полурешеток). Частичные порядки верхней и нижней полурешеток совпадают тогда и только тогда, когда выполняются законы поглощения.  $\square$

Доказательство. Сначала докажем необходимость. Если порядки совпадают, то заданное множество является одновременно и верхней и нижней полурешеткой, а, значит, является решеткой. Для решеток же законы поглощения выполняются [Мальцев, 1970], [Скорняков, 1986].

Докажем достаточность. Для этого нужно показать, что для  $\forall a, b$  выполняется эквивалентность:  $a + b = b \Leftrightarrow ab = a$ .

Допустим, что равенство  $a + b = b$  выполняется. Тогда  $ab = a(a + b)$ . По закону поглощения  $a(a + b) = a$ , поэтому  $ab = a$ .

Аналогично сделаем допущение, что  $ab = a$ . В этом случае  $a + b = ab + b$ . По закону поглощения  $ab + b = b$ , значит,  $a + b = b$ .  $\square$

Естественно, этот общий результат применим к построению решетки мультимножеств. Для этого надо убедиться в выполнении законов поглощения:  $\alpha \cap_{\text{All}} (\alpha \cup_{\text{All}} \beta) = \alpha$  и  $\alpha \cup_{\text{All}} (\alpha \cap_{\text{All}} \beta) = \alpha$ , что делается непосредственно (и, в свою очередь, вытекает из выполнения законов поглощения для теоретико-числовых функций  $\max$ ,  $\min$ ).

### Построение полной решетки мультимножеств

Приведем более сильные результаты о структуре введенного частично упорядоченного семейства мультимножеств.

**Предложение 1** (структура ч.у.м. мультимножеств). Выполняются следующие утверждения:

- 1) пустое мультимножество  $\emptyset_m$  (его характеристической функцией является константная функция, всюду равная нулю) – наименьший элемент в  $\langle M, \leq \rangle$ ;
- 2)  $\inf \mu = \alpha$ , для произвольного непустого множества<sup>1</sup> мультимножеств  $\mu \subseteq M$ ; здесь характеристическая функция мультимножества  $\alpha$  задается выражением  $\chi_{\alpha}(d) = \min_{\beta \in \mu} \chi_{\beta}(d)$ ;
- 3) для произвольного множества мультимножеств  $\mu$ : супремум  $\mu$  существует  $\Leftrightarrow \mu$  ограничено сверху;
- 4)  $\sup \mu = \alpha$ , где  $\mu$  – произвольное множество мультимножеств, имеющее точную верхнюю грань, а характеристическая функция мультимножества  $\alpha$  задается выражением  $\chi_{\alpha}(d) = \max_{\beta \in \mu} \chi_{\beta}(d)$ .  $\square$

Доказательство. Рассмотрим отдельно каждое из сформулированных выше утверждений.

- 1) Наименьшим элементом ч.у.м.  $\langle N, \leq \rangle$  является число 0. Соответственно, наименьшим элементом ч.у.м. характеристических функций  $\langle \{\chi_{\alpha} \mid \alpha \in M\}, \leq \rangle$  будет функция, всюду равная

<sup>1</sup> Отметим, что инфимум пустого множества мультимножеств не существует, так как в рассматриваемом частично упорядоченном множестве мультимножеств не существует наибольшего элемента.

- нулю. Исходя из того, что ч.у.м.  $\langle \{\chi_\alpha \mid \alpha \in M\}, \triangleleft \rangle$  изоморфно ч.у.м.  $\langle M, \preceq \rangle$ , получим, что наименьшим элементом второго будет пустое мультимножество  $\emptyset_m$ .
- 2) Доказательство второго утверждения вытекает из известной формулы нахождения инфимума множества функций (см., например, [Буй, 2002] или [Бурбаки, 1965]) и того факта, что ч.у.м.  $\langle N, \preceq \rangle$  является вполне упорядоченным.
  - 3) Супремум подмножества  $L$  ч.у.м.  $\langle N, \preceq \rangle$  существует тогда и только тогда, когда множество  $L$  – конечно (ограничено). Перенесем этот результат на ч.у.м. характеристических функций, а, следовательно, и на ч.у.м. мультимножеств (так как они изоморфны). Тогда получим, что супремум произвольного подмножества  $\mu$  ч.у.м.  $\langle M, \preceq \rangle$  существует, тогда и только тогда, когда это множество ограничено сверху.
  - 4) Доказательство этого утверждения также вытекает из известной формулы нахождения супремума множества функций (см., например, [Буй, 2002] или [Бурбаки, 1965]) при условии, что супремум существует.  $\square$

Таким образом, согласно предложению 1, ч.у.м.  $\langle M, \preceq \rangle$  имеет наименьший элемент (пустое мультимножество –  $\emptyset_m$ ) и любое его подмножество, ограниченное сверху, имеет точную верхнюю грань. Значит, следуя терминологии работ [Davey, 1990], [Биркгоф, 1984], ч.у.м.  $\langle M, \preceq \rangle$  является одновременно условно полным множеством и полной полурешеткой (complete semilattice). Таким образом, установлена теорема 3.

**Теорема 3.** Ч.у.м.  $\langle M, \preceq \rangle$  является условно полным множеством и полной полурешеткой, при этом точные грани находятся согласно формул предложения 1.  $\square$

Доказательство вытекает из определений условно полного множества и полной полурешетки, а также предложения 1.  $\square$

Пополним частично упорядоченное множество  $\langle M, \preceq \rangle$  наибольшим элементом  $T$ ; полученное ч.у.м. обозначим через  $\langle M \cup \{T\}, \preceq \rangle$ .

**Следствие 1.** Ч.у.м.  $\langle M \cup \{T\}, \preceq \rangle$  является полной решеткой с наименьшим элементом  $\emptyset_m$  и наибольшим элементом  $T$ .  $\square$

Доказательство проводится очевидным образом; кроме того, можно использовать общий результат теории решеток [Биркгоф, 1984].  $\square$

Ч.у.м.  $\langle M, \preceq \rangle$  можно вложить и в другую полную решетку. Для этого расширим понятие мультимножества.

С этой целью пополним множество натуральных чисел без нуля  $N^+$  наибольшим элементом  $\infty$  и положим  $N_\infty^+ = N^+ \cup \{\infty\}$ . Под мультимножеством будем понимать функцию вида  $\alpha : U_\alpha \rightarrow N_\infty^+$ , семейство всех таких мультимножеств обозначим через  $M_\infty$ . Порядок на множестве  $N_\infty^+$  обозначим через  $\leq_\infty$ , тогда порядок на мультимножествах  $\preceq$  расширяется так:  $\alpha \preceq_\infty \beta \stackrel{def}{\Leftrightarrow} \forall d (\chi_\alpha(d) \leq_\infty \chi_\beta(d))$ ,  $\alpha, \beta \in M_\infty$ .

**Теорема 4.** Ч.у.м.  $\langle M_\infty, \preceq_\infty \rangle$  является полной решеткой с наименьшим элементом  $\emptyset_m$  и наибольшим элементом  $T_\infty$ , где  $T_\infty : D \rightarrow \{\infty\}$ ,  $T_\infty(d) = \infty$  для всех  $d \in D$ . Точные нижние грани находятся по формулам предложения 1, для точных верхних граней выполняется формула  $\sup \mu = \alpha$ , где характеристическая функция мультимножества  $\alpha$  такая –  $\chi_\alpha(d) = \sup_{\leq_\infty, \beta \in \mu} \chi_\beta(d)$ .  $\square$

Доказательство. Очевидно, что  $\langle N_\infty^+, \leq_\infty \rangle$  является полной решеткой. Остается применить тот хорошо известный факт, что прямое произведение полных решеток будет полной решеткой, и использовать уже

неоднократно упоминавшуюся формулу для нахождения точных граней подмножеств прямого произведения. □

---

### Выводы

В работе приведены базовые определения, касающиеся теории мультимножеств: мультимножества, его характеристической функции, отношения включения, операций объединения и пересечения. Построена решетка мультимножеств. Выяснено устройство ч.у.м. мультимножеств по отношению включения мультимножеств. Осуществлено вложение ч.у.м. мультимножеств в две полные решетки, при этом одна из этих полных решеток строится путем обобщения понятия мультимножества (допускается бесконечная кратность элементов основы). Отметим, что такие обобщенные мультимножества необходимо рассматривать при задании денотационной семантики рекурсивных запросов в SQL-подобных языках [Буй, Поляков, 2010] (таким мультимножествам соответствуют таблицы с бесконечным числом строк); эта проблематика, однако, требует отдельного рассмотрения.

---

### Список литературы

- [Blizard, 1989] W. Blizard. The Development of Multiset Theory // Notre Dame Journal of Formal Logic. – 1989. – Vol. 30, No. 1. – P. 36-66.
- [Davey, 1990] B.A. Davey, H.A. Priestly. Introduction to Lattice and Order. Cambridge: Cambridge University Press, 1990. – 248 p.
- [Lamperti, 2001] G. Lamperti, M. Melchiori, M. Zanella. On Multisets in Database Systems // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 147-215.
- [Singh, 2007] D. Singh, A.M. Ibrahim, T. Yohanna, J.N. Singh. An Overview of the Applications of Multisets // Novi Sad Journal of Mathematics. – 2007. – Vol. 37, No. 2. – P. 73-92.
- [Syropoulos, 2001] A. Syropoulos. Mathematics of Multisets // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 347-358.
- [Биркгоф, 1984] Г. Биркгоф. Теория решеток. – Москва: Наука, 1984. – 564 с.
- [Буй, 2002] Д.Б. Буй. Теорія програмних алгебр композиційного типу та її застосування: дисертація доктора фізико-математичних наук: 01.05.03 – математичне та програмне забезпечення обчислювальних машин та систем. – Київ, 2002. – 365 с.
- [Буй, Богатирьова 2010] Д.Б. Буй, Ю.О. Богатирьова. Сучасний стан теорії мультимножин // Вісник Київського національного університету імені Тараса Шевченка. Сер. фіз.-мат. науки. – 2010. Вип. 1. – С. 51-58.
- [Буй, Поляков, 2010] Д.Б. Буй, С.А. Поляков. Композиційна семантика рекурсивних запитів в SQL-подібних мовах // Вісник Київського національного університету імені Тараса Шевченка. Сер. фіз.-мат. науки. – 2010. Вип. 1. – С. 45-50.
- [Бурбаки, 1965] Н. Бурбаки. Теория множеств. – Москва: Мир, 1965. – 455 с.
- [Мальцев, 1970] А.И. Мальцев. Алгебраические системы. – Москва: Наука, 1970. – 392 с.
- [Петровский, 2002] А.Б. Петровский. Основные понятия теории мультимножеств. – Москва: Едиториал УРСС, 2002. – 80 с.
- [Редько, 2001] В.Н. Редько, Ю.И. Брона, Д.Б. Буй, С.А. Поляков. Реляційні бази даних: табличні алгебри та SQL-подібні мови. – Київ: Видавничий дім "Академперіодика", 2001. – 198 с.
- [Скорняков, 1986] Л.А. Скорняков. Элементы алгебры. – Москва: Наука, 1986. – 240 с.

---

### Информация об авторах

**Буй Дмитрий Борисович** – заведующий сектором проблем программирования, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: 01601, ул. Владимирская, 64, Киев, Украина; e-mail: buy@unicyb.kiev.ua

**Богатирёва Юлия Александровна** – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: 01601, ул. Владимирская, 64, Киев, Украина; e-mail: j\_bogatyreva@ukr.net

## ОБОБЩЕННОЕ ИСЧИСЛЕНИЕ НА ДОМЕНЕ

Дмитрий Буй, Ирина Глушко

**Аннотация:** В данной статье доказано, что обобщенная табличная алгебра, исчисление строк (кортежей) и исчисление на домене, пополненные произвольными предикатными и функциональными сигнатурами на универсальном домене, одинаково выразительны. Проведена редукция исчисления строк к исчислению на домене. Доказано, что при этом обобщении табличная алгебра остается не менее выразительной, чем исчисление на домене.

**Ключевые слова:** реляционные (табличные) базы данных, исчисление на домене, табличная алгебра, исчисление строк (кортежей).

**АСМ классификация ключевых слов:** H.2 Database Management (E.5) H.2.4 Systems – Relational databases

### Введение

В предыдущих работах [Буй, Глушко, 2010], [Buy, Glushko, 2010] проведено обобщение табличной алгебры, построенной на основе реляционных алгебр Кодда; обобщение состоит в том, что вместо таблиц рассматриваются пары, состоящие из таблиц и схем таблиц. Сделан первый шаг в доказательстве эквивалентности табличной (реляционной) алгебры Кодда и исчисления строк (кортежей), пополненного произвольными предикатными и функциональными сигнатурами на универсальном домене: доказано, что исчисление строк при этом обобщении не менее выразительно, чем табличная алгебра.

Задача данной работы показать, что обобщенная табличная алгебра, исчисление строк (кортежей) и исчисление на домене, пополненные произвольными предикатными и функциональными сигнатурами на универсальном домене, одинаково выразительны.

### Обобщенное исчисление на домене

М. Лакруа (M. Lacroix) и А. Пиротте (A. Pirotte) [Lacroix, Pirotte, 1977] предложили альтернативную к исчислению кортежей версию исчисления – реляционное исчисление с переменными на доменах. Различие их в том, что вместо переменных строк рассматриваются переменные значения из домена, которые представляют компоненты строки. Кроме того, исчисление на домене поддерживает условие принадлежности (membership condition) [Дейт, 2005], [Хомоненко, Цыганков, Мальцев, 2004]. В общем виде условие принадлежности записывается как  $\langle t(\langle A_1, d_1 \rangle, \langle A_2, d_2 \rangle, \dots), R \rangle$ , где  $R$  – схема,  $A_i$  – атрибут таблицы  $t$ , а  $d_i$  – переменная домена или литерал (константа). Проверяемое условие истинно тогда и только тогда, когда в таблице  $t$  существует строка, которая имеет указанные значения из универсального домена  $D$  для указанных атрибутов. Причем порядок атрибутов в таблице фиксированный.

Все неопределенные в статье понятия и обозначение понимаем в смысле [Редько, Брона, Буй, Поляков, 2001].

Введем множество так называемых разрешенных формул исчисления на домене при использовании:

- множества атрибутов  $A$  и универсального домена  $D$ ;
- множества предметных констант  $d_1, d_2, \dots$ ;
- множества предметных переменных  $x_1, x_2, \dots$ ;
- множества функциональных символов  $f_1^{n_1}, f_2^{n_2}, \dots$ ;
- множества предикатных символов  $p_1^{m_1}, p_2^{m_2}, \dots$ .



Областью интерпретации предметных констант и предметных переменных есть универсальный домен  $D$ . Применяем  $x$  как синтаксическую переменную, областью изменения которой являются переменные;  $f(p)$  как синтаксическую переменную, областью изменения которой являются функциональные (соответственно предикатные) символы;  $d$  как синтаксическую переменную, областью изменения которой являются константы; наконец,  $\mathcal{A}$  как синтаксическую переменную, областью изменения которой являются атрибуты.

Следующие выражения являются термами (индукция по длине термов):

- a) всякая предметная константа или предметная переменная есть терм;
- b) если  $u_1, \dots, u_n$  – термы,  $f$  –  $n$ -арный функциональный символ, то  $f(u_1, \dots, u_n)$  – терм;
- c) выражение является термом только в том случае, если это следует из правил a), b).

Будем применять  $u$  как синтаксическую переменную, областью изменения которой являются термы.

Атомарные формулы (атомы) имеют вид:

- a1. пусть  $\langle t, R \rangle$  – таблица схемы  $R = \{A_1, \dots, A_n\} \subseteq A$ . Тогда  $t(a_1, \dots, a_n)$  – атом, где  $a_i$  – термы,  $i = 1, \dots, n$ ;
- a2. пусть  $u_1, \dots, u_m$  – термы, а  $p$  –  $m$ -арный предикат на универсальном домене  $D$ . Тогда  $p(u_1, \dots, u_m)$  – атом.

Используем логические связки  $\neg, \wedge, \vee$ , кванторы  $\exists, \forall$  и скобки  $(, )$  для построения формул из атомов.

- f1. Каждый атом – формула
- f2. Если  $P$  – формула, то  $\neg P$  – формула.
- f3. Если  $P$  и  $Q$  – формулы, то  $P \wedge Q, P \vee Q$  – формулы.
- f4. Пусть  $x$  – переменная,  $P$  – формула,  $\mathcal{A}$  – атрибут из множества  $A$ , тогда  $\exists x (\mathcal{A})P$  – формула.
- f5. Пусть  $x$  – переменная,  $P$  – формула,  $\neg \mathcal{A}$  атрибут из множества  $A$ , тогда  $\forall x (\mathcal{A})P$  – формула.
- f6. Если  $P$  – формула, то  $(P)$  – формула.
- f7. Других формул нет.

Понятия свободных и связанных переменных на домене и область действия связанной переменной определяются в исчислении на домене так же, как и в исчислении строк. Схема переменной  $x$  в формуле  $P$   $scheme(x, P)$  – это или универсальный домен  $D$ , или неопределенная величина. Формально определять понятие разрешенной формулы не будем. Как и в исчислении строк формула разрешена, если выполняется требование согласования схем переменной в подформулах и требование того, что переменная, связанная квантором, должна входить свободно в формулу, которая следует за квантором.

Выражение исчисления с переменными на домене имеет вид:  $\{x_1, \dots, x_n \mid P(x_1, \dots, x_n)\}$ , где

1. формула  $P$  – разрешена, а  $x_1, \dots, x_n$  – все свободные переменные, входящие в формулу  $P$ ;
2.  $R = \{A_1, \dots, A_n\}, R \subseteq A$  – схема, порядок атрибутов фиксированный;
3.  $scheme(x_i, P) = D, i = 1, \dots, n$ .

Пусть  $P(x)$  – разрешенная формула,  $\mathcal{A}$  – атрибут из множества  $A$  и если  $scheme(x, P)$  определена, то  $scheme(x, P) = D$ . В результате подстановки константы  $d$  из универсального домена  $D$  вместо переменной  $x$  в формулу  $P$  получим формулу обозначенную  $P(d/x)$ . Определим сначала истинностные значения атомов:

a1. пусть переменная  $x$  в подформуле  $t(x)$  свободна в  $P$ ,  $R = \{A\}$  – схема таблицы  $\langle t, R \rangle$ . Атом  $t(x)$  приобретает значение истина, при подстановке константы  $d \in D$  вместо  $x$ , если  $\exists s(s \in t \wedge \langle A, d \rangle \in s)$ , где  $s$  – строка таблицы  $\langle t, R \rangle$ , иначе атом  $t(x)$  приобретает значение ложь;

a2. пусть переменная  $x$  в подформуле  $p(u_1, \dots, u_m)$  свободна в  $P$ , тогда при подстановке константы  $d \in D$  вместо переменной  $x$ , заменим соответствующую предметную переменную на  $d$ . Атом  $p(u_1, \dots, u_m)$ , где  $u_i$  – термы (предметные константы), приобретает значение истина, если предикат  $p$  истинный на соответственных значениях, иначе атом приобретает значение ложь.

Пусть формула  $P$  разрешенная формула без свободных переменных. Интерпретация формулы  $P$  определяется следующим образом.

f2. Если  $P = \neg G$ , то в  $G$  нет свободных переменных. Формула  $P$  истинная, когда  $G$  ложная, и ложная, когда формула  $G$  истинная.

f3. Если  $P = G \wedge Q$  или  $P = G \vee Q$ , то в  $G$  и  $Q$  нет свободных переменных. Если  $P = G \wedge Q$ , то формула  $P$  истинная тогда, когда  $G$  и  $Q$  обе истинные, в противном случае формула  $P$  ложная. Если  $P = G \vee Q$ , то формула  $P$  ложная тогда, когда  $G$  и  $Q$  обе ложные, в противном случае формула  $P$  истинная.

f4. Если  $P = \exists x(A)G$ , то  $x$  – единственная свободно входящая в формулу  $G$  переменная. Формула  $P$  истинная, если существует константа  $d \in D$ , такая, что формула  $G(d/x)$  – истинная, иначе формула  $P$  ложная.

f5. Если  $P = \forall x(A)G$ , то  $x$  – единственная свободно входящая в формулу  $G$  переменная. Формула  $P$  истинная, если для каждой константы  $d \in D$  формула  $G(d/x)$  истинная, иначе формула  $P$  ложная.

f6. Если  $P = (G)$ , то формула  $P$  истинная, если формула  $G$  истинная, и ложная, если формула  $G$  ложная.

Пусть  $E = \{x_1, \dots, x_n \mid P(x_1, \dots, x_n)\}$  – выражение исчисления на домене. Значением выражения  $E$  назовем таблицу схемы  $R = \{A_1, \dots, A_n\}$ , содержащую все строки  $s \in S(R)$ ,  $s = \{\langle A_1, d_1 \rangle, \dots, \langle A_n, d_n \rangle\}$ , такие, что формула  $P(d_1/x_1, \dots, d_n/x_n)$  истинная.

### Сведение исчисления строк к исчислению на домене

Пусть  $F = \{y(R) \mid P(y)\}$  – выражение исчисления строк, переменная  $y$  – единственная переменная, входящая в формулу  $P$  свободно, а другие переменные связаны не более, чем одним квантором,  $R = \{A_1, \dots, A_n\}$  – схема. Применяем  $y$  как синтаксическую переменную, областью изменения которой являются переменные исчисления строк. Областью интерпретации предметных переменных исчисления строк есть множество всех строк. Построим выражение исчисления на домене, эквивалентное заданному выражению исчисления строк  $F$ . Каждая переменная  $z$  исчисления строк, входящая в формулу  $P$ , соответствует некоторой схеме  $R_2$  таблицы, причем либо  $z$  входит в  $\exists z(R_2)$  или  $\forall z(R_2)$ , либо  $z = y$ , а  $R_2 = R$ . Пусть  $R_2 = \{A_1, \dots, A_m\}$ . Выполним следующие замены.

- Переменную  $z$  исчисления строк заменим  $m$  переменными  $z_1, \dots, z_m$  исчисления на домене.
- Каждый атом  $t(z)$  исчисления строк заменим на  $t(z_1, \dots, z_m)$ .

- Каждый атом  $p(\mathbf{v}_1, \dots, \mathbf{v}_m)$ , где  $\mathbf{v}_i$  – термы исчисления строк,  $i = 1, \dots, m$ , заменим на  $p(\mathbf{u}_1, \dots, \mathbf{u}_m)$ , где  $\mathbf{u}_j$  – термы исчисления на домене,  $j = 1, \dots, m$ .
- Формулу с квантором  $\exists \mathbf{z}(R_2)G$  преобразуем в  $\exists \mathbf{z}_1(A_1) \dots \exists \mathbf{z}_m(A_m)G$ .
- Формулу с квантором  $\forall \mathbf{z}(R_2)G$  преобразуем в  $\forall \mathbf{z}_1(A_1) \dots \forall \mathbf{z}_m(A_m)G$ .
- Если  $\mathbf{z} = \mathbf{y}$ , а  $R_2 = R$ , то заменим  $\mathbf{y}(R)$  в начале выражения  $F$  на  $\mathbf{z}_1, \dots, \mathbf{z}_m$ .

В результате проведенных замен получим, очевидно, выражение исчисления с переменными на домене, эквивалентное заданному выражению исчисления строк. Поэтому можем сформулировать следующую теорему.

**Теорема.** Если  $F$  – произвольное разрешенное выражение исчисления строк, то можно эффективно (и равномерно) построить эквивалентное ему выражение  $E$  исчисления на домене.

### Сведение исчисления на домене к табличной алгебре

Рассматриваем два множества:  $\mathbf{A}$  – множество атрибутов и  $\mathbf{D}$  – универсальный домен. Под табличной алгеброй понимаем алгебру  $\langle \mathbf{T}, \Omega_{P, \Xi} \rangle$ , где  $\mathbf{T}$  – множество всех таблиц,  $\Omega_{P, \Xi} = \{ \cup_R, \cap_R, \setminus_R, \sigma_{p, R}, \pi_{X, R}, \otimes_{R_1, R_2}, \div_{R_2}, Rt_{\xi, R}, \sim_R \}_{p \in P, \xi \in \Xi, X \subseteq R, R_1, R_2 \subseteq A}$  – сигнатура,  $P, \Xi$  – множества параметров. Выражением табличной алгебры называется любое выражение, построенное из таблиц множества  $\mathbf{T}$  при использовании операций множества  $\Omega_{P, \Xi}$ .

**Теорема.** Для каждого выражения исчисления на домене  $E$  можно эффективно (и равномерно) построить эквивалентное ему выражение  $F$  табличной алгебры.

**Доказательство.** Пусть  $E = \{ \mathbf{x}_1, \dots, \mathbf{x}_n \mid P(\mathbf{x}_1, \dots, \mathbf{x}_n) \}$  – выражение исчисления на домене, значением которого есть таблица схемы  $R = \{ A_1, \dots, A_n \}$ . Построим для каждой подформулы  $G$  формулы  $P$  эквивалентное алгебраическое выражение  $F_G$ . Докажем индукцией по числу операторов в подформуле  $G$  из  $P$ , что если  $G$  содержит свободные переменные на универсальном домене  $\mathbf{y}_1, \dots, \mathbf{y}_m$ , то для выражения  $\{ \mathbf{y}_1, \dots, \mathbf{y}_m \mid G(\mathbf{y}_1, \dots, \mathbf{y}_m) \}$  существует эквивалентное выражение табличной алгебры  $F_G$ . Тогда, когда  $G$  есть само  $P$ , получим алгебраическое выражение для  $\{ \mathbf{x}_1, \dots, \mathbf{x}_n \mid P(\mathbf{x}_1, \dots, \mathbf{x}_n) \}$ . Предполагается, что  $\mathbf{y}_1, \dots, \mathbf{y}_m$  – все свободные переменные в  $G$ , и значением выражения  $\{ \mathbf{y}_1, \dots, \mathbf{y}_m \mid G(\mathbf{y}_1, \dots, \mathbf{y}_m) \}$  является таблица схемы  $R_G = \{ B_1, \dots, B_m \} \subseteq \mathbf{A}$ .

Заменим связанные переменные в формуле  $P$  так, чтобы ни одна переменная не входила одновременно и свободно, и связано в  $P$ , а также не связывалась квантором в двух разных местах. Каждая переменная ассоциирована теперь с атрибутом, либо входящим в квантор  $\forall \mathbf{x}(\mathcal{A})$  или  $\exists \mathbf{x}(\mathcal{A})$ , когда переменная связана в формуле  $P$ , либо сопоставленным с переменной, стоящей слева от вертикальной черты.

Для каждой формулы  $P$  исчисления на домене можно выписать алгебраическое выражение со значением  $\mathbf{D}$  (значением данного алгебраического выражения есть одноатрибутная таблица  $\langle t, \{A\} \rangle$ , содержащая все строки  $s = \langle \langle A, d \rangle \rangle$ ,  $d \in \mathbf{D}$  – все элементы множества  $\mathbf{D}$ ). Например,  $\langle \langle A, d \rangle \rangle \cup \overline{\langle \langle A, d \rangle \rangle}$ , где  $d$  – произвольный элемент из  $\mathbf{D}$ . Далее  $[D]$  будет обозначать алгебраическое выражение со значением  $\mathbf{D}$ .

База индукции (нет операторов, т.е. логических связей и кванторов). Подформула  $G$  – атом вида  $p(\mathbf{u}_1, \dots, \mathbf{u}_m)$  или  $t(\mathbf{a}_1, \dots, \mathbf{a}_n)$ .

1) Пусть  $G$  имеет вид  $p(u_1, \dots, u_m)$ , где  $u_i$  – термы исчисления на домене. Положим  $F_G$  равным  $\sigma_{\tilde{p}}([D]_1 \otimes_{R_1, R_2} \dots \otimes_{R_{k-1}, R_k} [D]_k)$ , где  $R_i$  – одноатрибутные схемы таблиц, которые есть значениями алгебраических выражений  $[D]_i$ ,  $i = 1, \dots, k$ . Соединение в данном случае является декартовым произведением. А предикат-параметр селекции задан как  $\tilde{p}(s) = T \Leftrightarrow p(s(A_1), \dots, s(A_m)) = T$ ,  $s \in S(R)$ , где  $p$  –  $m$ -арный сигнатурный предикатный символ.  $\square$

2) Пусть  $G$  имеет вид  $t(a_1, \dots, a_n)$ , где  $a_i$  – константа или переменная на универсальном домене  $D$ . Пусть  $R = \{C_1, \dots, C_k\}$  – схема таблицы  $\langle t, R \rangle$ . Положим алгебраическое выражение  $F_G$  равным  $\pi_X(\sigma_{\tilde{p}}(t))$ , где  $\tilde{p}$  – предикат-параметр селекции, который является конъюнкцией сравнений  $C_i = a_i$  по всех  $i$ , для которых  $a_i$  – константа;  $X$  – множество атрибутов  $\{C_j \mid a_j \text{ – переменная}\}$ .  $\square$

Шаг индукции. Предположим, что подформула  $G$  содержит по крайней мере один оператор и что гипотеза индукции справедлива для всех подформул формулы  $P$ , имеющих меньше операторов, чем  $G$ .

Случай 1.  $G = \neg Q$ . Пусть  $F_Q$  – алгебраическое выражение для  $Q$ . Положим  $F_G = \neg F_Q$ .  $\square$

Случай 2.  $G = Q \vee Q'$ . Пусть свободные переменные из  $Q$  – это  $z_1, \dots, z_k, v_1, \dots, v_p$ , а из  $Q'$  – это  $z_1, \dots, z_k, w_1, \dots, w_q$ , где переменные  $v_1, \dots, v_p$  и  $w_1, \dots, w_q$  попарно не совпадают. Пусть  $F_Q$  и  $F_{Q'}$  – алгебраические выражения для  $Q$  и  $Q'$  соответственно, причем  $C_i$ ,  $i = 1, \dots, p$  – атрибуты ассоциированные с переменными  $v_1, \dots, v_p$ , а  $K_j$ ,  $j = 1, \dots, q$  – атрибуты ассоциированные с переменными  $w_1, \dots, w_q$ . Положим  $F_1 = F_Q \otimes_{R_Q, \{K_1\}} [D]_1 \otimes_{\{K_1\}, \{K_2\}} \dots \otimes_{\{K_{q-1}\}, \{K_q\}} [D]_q$ , а

$F_2 = F_{Q'} \otimes_{R_{Q'}, \{C_1\}} [D]_1 \otimes_{\{C_1\}, \{C_2\}} \dots \otimes_{\{C_{p-1}\}, \{C_p\}} [D]_p$ , где  $R_Q$  и  $R_{Q'}$  – схемы таблиц, которые являются значениями алгебраических выражений  $F_Q$  и  $F_{Q'}$  соответственно, а  $C_i$  и  $K_j$  – одноатрибутные схемы таблиц, которые являются значениями алгебраических выражений  $[D]_i$  и  $[D]_j$  соответственно,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$ . Обозначим схемы таблиц, которые являются значениями алгебраических выражений  $F_1$  и  $F_2$  как  $R_{F_1}$  и  $R_{F_2}$  соответственно. Для соответствующим образом построенных выражений  $F_1$  и  $F_2$ , их схемы  $R_{F_1}$  и  $R_{F_2}$  равны. Поэтому положим  $F_G = F_1 \cup_{R_{F_1}} F_2$ .  $\square$

Случай 3.  $G = Q \wedge Q'$ . Это выражение исчисления на домене можно выразить через операции отрицания и дизъюнкции. Следуя закону де Моргана  $G = \neg(\neg Q \vee \neg Q')$ .  $\square$

Случай 4.  $G = \exists x(A)Q$ . Пусть  $F_Q$  – алгебраическое выражение для  $Q$ . Положим  $F_G$  равным  $\pi_{X \setminus A}(F_Q)$ , где  $X$  – схема таблицы, которая является значением алгебраического выражения  $F_Q$ .  $\square$

Случай 5.  $G = \forall x(A)Q$ . Пусть  $F_Q$  – алгебраическое выражение для  $Q$ . Положим  $F_G$  равным  $F_Q \div_{R_1, R_2} [D]$ , где  $R_1$  – схема таблицы, которая является значением алгебраического выражения  $F_Q$ ,  $R_2$  – схема таблицы, которая является значением алгебраического выражения  $[D]$ .  $\square$

## Выводы

Классическое исчисления на домене пополнено произвольными предикатными и функциональными сигнатурами на универсальном домене (в то время как обычно рассматривают лишь бинарные предикаты, а функциональная сигнатура вообще пуста). Определено синтаксис термов, атомов и формул исчисления на домене; выделен класс разрешенных формул. Проведена редукция исчисления строк к исчислению на домене. Доказано, что при этом обобщении табличная алгебра остается не менее

выразительной, чем исчисление на домене. Учитывая, что ранее доказано, что исчисление кортежей не менее выразительно, чем табличная алгебра, можно сделать вывод, что табличная алгебра, исчисление строк и исчисление на домене одинаково выразительны.

---

### Література

---

[Буй, Глушко, 2010] Д.Б. Буй, І.М. Глушко. Узагальнене числення рядків // Наукові записки НаУКМА. Серія: Комп'ютерні науки. Київ, 2010.

[Buy, Glushko, 2010] Generalized table algebra, generalized tuple calculus and theirs equivalence // Proceeding of the CSE'2010 International Scientific Conference on Computer Science and Engineering. – Košice - Stará Ľubovňa, Slovakia. – 2010, September 20-22.

[Lacroix, Pirotte, 1977] M. Lacroix, A. Pirotte. Domain-oriented relational languages / M. Lacroix, A. Pirotte. In: Proc. 3rd Int. Conf on Very Large Data Bases. Tokyo, October, 1977. – P. 370-378.

[Дейт, 2005] К. Дж. Дейт. Введение в системы баз данных: [8-е изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2005.

[Хомоненко, Цыганков, Мальцев, 2004] А. Д. Хомоненко, В. М. Цыганков, М. Г. Мальцев. Базы данных. Санкт-Петербург: КОРОНА принт, 2004.

[Редько, Брона, Буй, Поляков, 2001] Реляційні бази даних: табличні алгебри та SQL-подібні мови / В. Н. Редько, Ю. Й. Брона, Д. Б. Буй, С. А. Поляков. Київ: Видавничий дім «Академперіодика», 2001.

[Мейєр, 1987] Д. Мейєр. Теория реляционных баз данных. Москва: Мир, 1987.

[Ульман, 1983] Дж. Ульман. Основы систем баз данных. Москва: Финансы и статистика, 1983.

[Коннолли, Бегг, 2003] Т. Коннолли, К. Бегг. Базы данных. Проектирование, реализация и сопровождение. Теория и практика: [3-изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2003.

---

### Информация об авторах

---

**Дмитрий Буй** – доктор физико-математических наук, старший научный сотрудник, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: [buy@unicyb.kiev.ua](mailto:buy@unicyb.kiev.ua).

**Ирина Глушко** – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: [glushkoim@gmail.com](mailto:glushkoim@gmail.com).

## DISTURBANCE OF STATISTICAL STABILITY

Igor Gorban

**Abstract:** *The survey of author's researches devoted to disturbance of statistical stability of physical quantities and processes is presented. Concepts of statistical stability of random sequences and random processes are formalized. Parameters characterizing their statistical stability in finite observation interval are proposed. It is found that statistical unstable processes are especial class of non-stationary processes. Statistical stability disturbance can be explained by low frequency cyclical fluctuations of expectation. It has been researched statistical stability of a number of physical quantities and processes, in particular line voltage, height of sea waves and their periods, Earth magnetic field fluctuation, and currency rate. It has been found that for all researched quantities and processes, there is statistical stability disturbance on large observation intervals. Obtained results corroborate the hypotheses that are in base of new physico-mathematical theory of hyper-random phenomena: real physical events, quantities, processes, and fields are not statistical stable and they can be adequately described by hyper-random models taking into consideration their statistical instability.*

**Keywords:** *statistical instability, theory of hyper-random phenomena, uncertainty, probability.*

**ACM Classification Keywords:** *G.3 Probability and Statistics*

---

### 1. Introduction

Physical basis of probability theory and mathematical statistics is statistical stability of event's frequencies. This fact examined and tested a lot of times by different scientists. It is well known, for instant, coin test led by Buffon and K. Pirson [Гнеденко, 1961]. Buffon flipped a coin 4040 times. K. Pirson led two test series. In the first series a coin was flipped by him 12000 times and in the second series – 24000 times. In all their tossing experiments the head frequencies equaled to near 0.5. Such stable result is typical for any statistical stable conditions.

In statistics and probability theory stability plays very important role. This circumstance was marked by a number of scientists, beginning from Jakob Bernoulli [Bernoulli, 1713]. R. von Mises proposed even to define [Mises, 1964] the concept of probability on the base of event's frequency in fixed conditions. Although Mises's proposition was not supported by the most mathematics and now Kolmogorov's set-theoretic definition of the random event [Колмогоров, 1936, ISO, 2006] is used mainly, importance of statistical stability is not decreased, especially in sphere of applications.

For real physical phenomena (events, quantities, processes, and fields) it is not simple to determine correctly Kolmogorov's probability measure. This fact was marked in many works, for instance, in [Колмогоров, 1986, Скороход, 1990, Тутубалин, 1972, Alimov, Kravtsov, 1992]. The main cause is in difficulty and often impossibility to stabilize statistical conditions.

This stimulated the development of new theories, such as fuzzy logic [Zadeh, Kacprzyk, 1992], neural network [Hagan, Demuth, Beale, 1996], chaotic dynamical systems [Crownover, 1995], interval data [Шокин, 1981, Алефельд, Херцбергер, 1987, Shary, 2002, Kreinovich, Berleant, Ferson, Lodwick, 2005, Ferson, Kreinovichy, Ginzburg, Myers, 2003], and others theories.

Author's apprehension in statistic stability of real mass phenomena motivated him to develop physico-mathematical theory of hyper-random phenomena [Горбань, 2007, Gorban, 2008, Gorban, 2009] oriented to description of statistically unstable physical events, quantities, processes, and fields.

The theory of hyper-random phenomena includes two components: mathematical and physical ones. Don't discussing now mathematical part of it, mark that its physical part is based on the hypothesis of existence in real physical world statistically unstable physical phenomena and the hypothesis of hyper-random setting up of the world, essence of which is that statistical unstable phenomena (adequately described by hyper-random models) are mass ones.

Now the basis of new theory is developed in different directions, however, for this time the main question remains disputable: are real phenomena statistically unstable or not?

To obtain the answer to the question a number of experimental researches devoted to disturbance of statistical stability of real physical quantities and processes were led by the author [Горбань, 2010 (1), Горбань, 2010 (2), Горбань, 2010 (3), Горбань, 2010 (4)].

The aim of the article is to present generalized results of these experimental researches.

## 2. Experimental research of line voltage

Research of line voltage is led by computer and simple linkage (consists of transformer and voltage divider). Input voltage was discretized with frequency 5 kHz. Voltage effective values were calculated on the base of every 1024 units and were written in memory. Record was led by session during two months with pause in some days. Duration of every record was 60 hours. For this session time, near  $N = 2^{20} \approx 1$  million of effective voltage units were recorded.

Line voltage was changed continuously. In different sessions changes had different character. Time dependences of line voltage for two sessions and according cumulative averages are presented in Fig. 1.

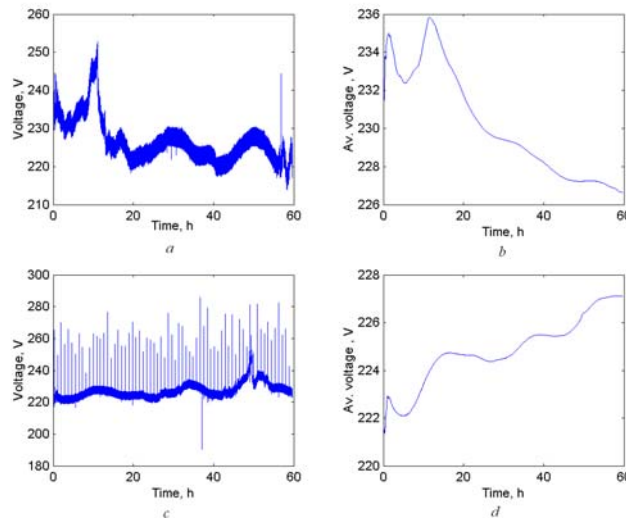


Fig. 1. Time dependences of line voltage for two sessions (in hours) (a, c) and according cumulative averages (b, d)

Analyze of a number records showed that all cumulative averages are not decayed (Fig. 1 b, d). This strange on the first view fact, sharply contrasts with well known theoretical results, demonstrating time decay of cumulative averages when the observation time is increased. Such decay is typical, for instance, for white Gaussian noise (model 1) and harmonic oscillation (model 2) presented in Fig. 2.

To clear causes of such difference let us appeal to the theory.

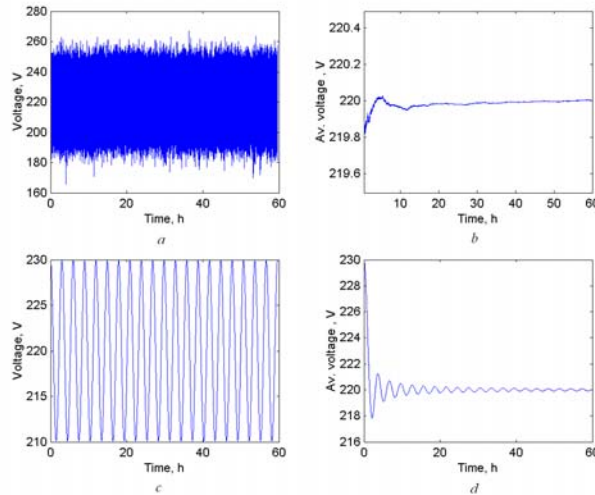


Fig. 2. White Gaussian noise (model 1) (a), harmonic oscillation (model 2) (c), and according cumulative averages (b, d)

### 3. Some theory questions

Sequence  $X_1, X_2, \dots$  of random variables (random sample) will be called a statistically stable one [Горбань, 2010 (1)], if expectation of sample variance  $\bar{D}_{Y_N} = \frac{1}{N} \sum_{n=1}^N (Y_n - \bar{m}_{Y_N})^2$  of average  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$  ( $n = \overline{1, N}$ ) tends to zero

when sample size  $N$  tends to infinity where  $\bar{m}_{Y_N} = \frac{1}{N} \sum_{n=1}^N Y_n$  is sample mean of average. Sequences that don't accord to this clause will be called statistically unstable ones.

The type of convergence is not essential here but to obtain necessary mathematical rigor we shall mean probability convergence.

It is well known Chebyshev theorem that presents the law of large numbers states [Гнеденко, 1961]: average  $Y_N = \frac{1}{N} \sum_{n=1}^N X_n$  of sequence  $X_1, X_2, \dots$  of independent in pair random variables, that have finite variances and

expectations  $m_{x_1}, m_{x_2}, \dots$ , tends in probability to mean  $m_{y_N} = \frac{1}{N} \sum_{n=1}^N m_{x_n}$  when sample size  $N$  tends to infinity.

Draw attention on one nuance unnoticed by the most: this theorem does not say about convergence of neither average  $Y_N$ , nor mean  $m_{y_N}$ . It states the convergence of these variables to each other, or otherwise, the convergence of their difference to zero. This means that average  $Y_N$  and mean  $m_{y_N}$  can have not the limit. They can fluctuate around the constant. In the process of such fluctuation they are changed without fail by synchronously manner.

It follows from Chebyshev theorem, that random sequence is statistically stable if and only if sample variance of mean  $m_{y_N}$  tends to zero when sample size  $N$  tends to infinity.

Random process will be called a statistically stable one, if expectation of the integral  $\frac{1}{T} \int_0^T (Y(t) - \bar{m}_{Y_T})^2 dt$  tends

to zero when  $T \rightarrow \infty$  where  $Y(t) = \frac{1}{t} \int_0^t X(t_1) dt_1$  is average of  $X(t_1)$  in the interval  $(0, t)$ ,  $\bar{m}_{Y_T} = \frac{1}{T} \int_0^T Y(t) dt$  is mean of average  $Y(t)$ . Random processes that don't accord to this clause will be called statistically unstable ones.



Mark, that determinate quantity  $x_0$  can be regarded in rough as a particular case of a random quantity that has in the distribution function a unit step in the point  $x_0$  [Горбань, 2007, Gorban, 2009]. Analogues, determinate function  $x_0(t)$  can be regarded in rough as a particular case of a random function with distribution function  $F(x;t) = \text{rect}(x - x_0(t))$ . Therefore statistically stable and statistically unstable concepts are suitable for sequence of determinate quantities and for sequence of determinate functions.

Examples of statistically stable sequences are uniform random sample with finite two first moments and also nonuniform random sample with finite two first moments, for which mean of expectations has a limit.

The sequence is statistically unstable one if mean of expectations has not a limit (for instance, fluctuates) when sample size tends to infinity. Numerical divergent sequence is statistically unstable one too.

Underline, for random processes, concept of nonstationarity and concept of statistical instability are not identical.

Stationary ergodic (in expectation) processes are statistically stable ones. Among non stationary processes there are as statistically stable as statistically unstable processes. Hence, statistically unstable processes are especial class of non stationary processes.

The fact of statistical stability or statistical instability of real sequence with finite sample size or the same fact of real process on the finite interval observation cannot be established in principle, because for establishing such facts the sequence or the process must be infinite ones.

However, concept of statistical stability on finite interval may be formalized. The bases for such formalization may be detection the tendency of stabilization of the average level or the tendency of stabilization of the mean of expectations when sample size increases. These tendencies may be the quality indicators of statistic stability of the process. Before going to quantity characteristics let us analyze possible causes led to disturbance of statistical stability.

#### 4. Possible causes led to disturbance of statistical stability of random processes

Any non stationary random process  $X(t)$  can be presented in the following form:

$$X(t) = m_{x(t)} + \dot{X}(t)$$

where  $m_{x(t)}$  is expectation of the process  $X(t)$  and  $\dot{X}(t)$  is random process with zero expectation.

Expectation  $m_{y(t)}$  of average  $Y(t)$  is determined by expectation  $m_{x(t)}$ :

$$m_{y(t)} = \frac{1}{t} \int_0^t m_{x(t_1)} dt_1.$$

Hence statistic stability of the process  $X(t)$  depends from particularities of expectation  $m_{x(t)}$ .

Examine random processes with different changes of expectation  $m_{x(t)}$ , in particular periodic, intermittent, and aperiodic types.

Let  $m_{x(t)}$  is a periodic function with period  $T$ . Then its Fourier expansion

$$m_{x(t)} = \sum_{k=-\infty}^{\infty} \hat{a}_k \exp\left(\frac{j2\pi}{T} kt\right) \quad (1)$$

where  $\hat{a}_k = a_k \exp(j\varphi_k)$  is complex expansion coefficient.

Then average expectation

$$m_{y(t)} = a_0 + \sum_{k=1}^{\infty} a_k \frac{\sin \pi tk/T}{\pi tk/T} \cos(\pi tk/T + \varphi_k). \quad (2)$$

It follows from expression (2) that variable part of average expectation is described by harmonic functions subsided on the low of  $\sin x / x$ . Subside speed of these functions is determined by value of the period  $T$ : it decreases with rising the period and tends to zero when  $T \rightarrow \infty$ .

Minimum subside speed has the first term of series (according to  $k = 1$ ). In average expectation (2) harmonics of high orders are suppressed. Suppress level increases with increasing of harmonic order.

Changing of average expectation is not significant if observation interval  $t$  essentially less than the period  $T$ . This is area of statistical stability. However, situation gradually changes when value  $t$  approaches to the period  $T$ . As follow from expression (2), average expectation is changed significantly in the interval  $t \in [0, T]$ . It means that in the interval  $t \in [0, T]$  statistical stability is disturbed.

Mark, essential changes of average expectation and disturbances of statistic stability may be registered in the observation intervals that larger than period  $T$  too, in particular, when amplitudes of high order harmonics are large and numbers of these harmonics are not high. Fig. 3 demonstrates these particularities (models 3, 4).

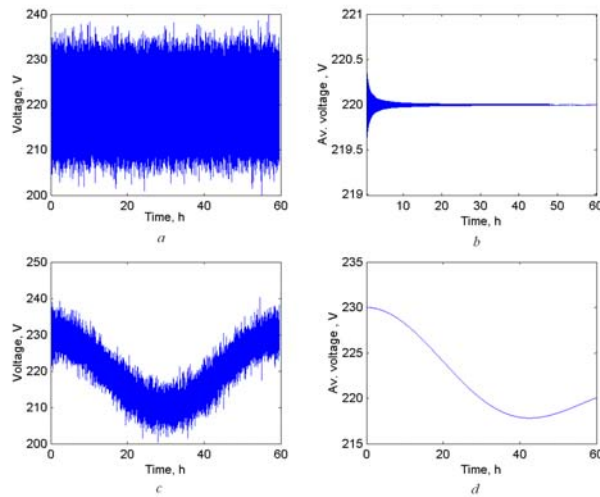


Fig. 3. Models of random processes with high (model 3) (a) and low (model 4) (c) oscillation frequency of expectation and according averages (b, d)

It follows from expression (2), that fluctuation of average expectation  $m_{y(t)}$  tends to zero when  $t \rightarrow \infty$ . This means that, although random process with periodic oscillation of expectation has in some observation intervals disturbances of statistical stability, it is statistically stable in infinite interval.

Let  $X(t)$  is a random process consists of  $Q$  processes  $X_q(t)$  ( $q = \overline{1, Q}$ ) with near equal levels and periodic expectations. Periods of these expectations are  $T_1, T_2, \dots, T_Q$ . The period  $T_{q+1}$  of each process  $X_{q+1}(t)$  is essentially larger than the period  $T_q$  of previous process  $X_q(t)$ .

In observation interval from zero to  $t$  that is essentially less than  $T_1$ , there are no essential changes of expectations and therefore the process  $X(t)$  is statistically stable, in practice. With approaching  $t$  to the period  $T_1$  the process  $X_1(t)$  (and therefore the process  $X(t)$ ) becomes more statistically unstable. With further rising of observation time statistical properties of the process  $X_1(t)$  appear and the process  $X(t)$  becomes similar to statistically stable one.

With approaching  $t$  to the period  $T_2$  the process  $X_2(t)$  becomes more statistically unstable again. This leads to disturbance of statistical stability of the process  $X(t)$ , and so on. When  $Q \rightarrow \infty$  and  $T_Q \rightarrow \infty$  interchange of statistically stable and unstable states reaches to infinite observation interval.

When  $T_{q+1} < 2T_q$  unstable areas are joined and process  $X(t)$  is statistically unstable in all observation interval  $[T_1, T_Q)$ .

Forming process of statistically unstable areas is illustrated by the models 5 and 6 (Fig. 4).

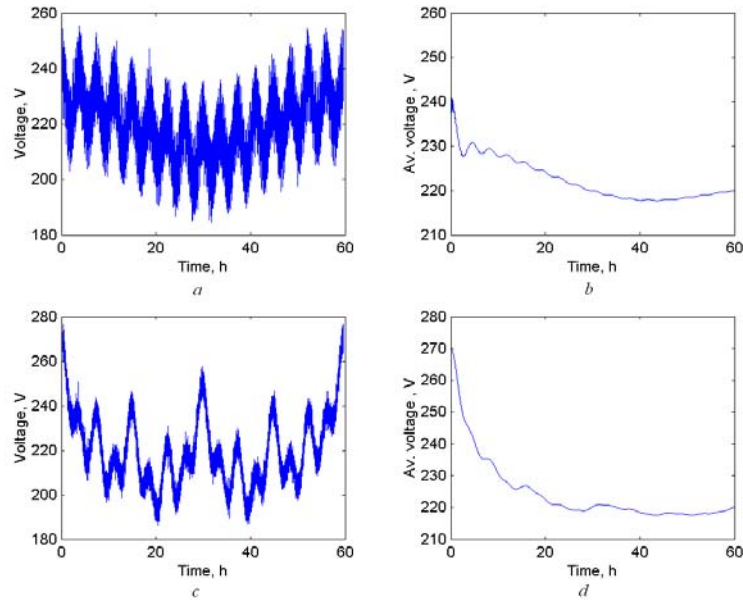


Fig. 4. Model of random process with expectation that has tree essentially different on frequency harmonics (model 5) (a), model of random process with five close to each other on frequency harmonics (model 6) (c), and according averages (b, d)

Described additive model explains interchanging of statistically stable and unstable states in real processes.

Random processes with intermittent changing of expectation are not interesting very much because distinguished samples create changes in average, however these changes are smoothed out fast (model 7, Fig. 5, a, b).

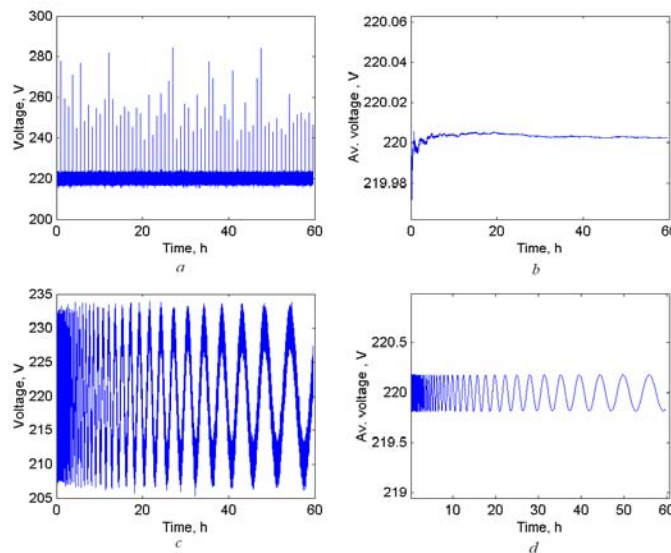


Fig. 5. Model of random process with intermittent change of expectation (model 7) (a), model of random process with aperiodic change of expectation (model 8) (c), and according averages (b, d)

Random processes with aperiodic changing of expectation are more interesting. Mark, process  $X(t)$  with expectation, which changes periodically with the period  $T$ , may be regarded in the observation interval  $t < T$  as aperiodic process. So, aperiodic processes may be statistically unstable.

If expectation  $m_{x(t)}$  may be presented as expansion in a Taylor series

$$m_{x(t)} = \sum_{k=0}^{\infty} a_k t^k, \quad (3)$$

then

$$m_{y(t)} = \sum_{k=0}^{\infty} \frac{a_k t^k}{k+1}. \quad (4)$$

It follows from expressions (3) and (4), if expectation  $m_{x(t)}$  is changed on low  $t^k$  expectation  $m_{y(t)}$  is changed on the same low too. So, if  $m_{x(t)} = t^k$ , the process  $X(t)$  is not statistically stable in any observation interval.

Mark, in compare with expansion (3), expansion (4) has addition coefficient  $(k+1)^{-1}$ . Therefore, in general, changing low of expectation  $m_{y(t)}$  doesn't repeat changing low of expectation  $m_{x(t)}$  and so, processes with aperiodic changing of expectations are not always statistically unstable ones.

The process  $X(t)$ , expectation of which is changed periodically in logarithmic scale with period  $T$ , has interesting property. In this case, expectation of the process

$$m_{x(t)} = \sum_{k=-\infty}^{\infty} \hat{a}_k \exp\left(\frac{j2\pi}{T} k \ln t\right).$$

According average expectation

$$m_{y(t)} = a_0 + 2 \sum_{k=1}^{\infty} \frac{a_k}{\sqrt{1 + 4\pi^2 k^2 / T^2}} \sin(2\pi k \ln t / T + \varphi_k + \arctg(T / 2\pi k)).$$

As follow from last expression, for the process  $X(t)$ , expectation of which is changed periodically in logarithmic scale, average expectation  $m_{y(t)}$  is described by harmonic non subsided functions. This mean that the process is not statistically stable in the interval  $[0, \infty)$ . Such process (model 8) and according average are presented in Fig. 5, c, d.

##### 5. Random processes in finite observation interval: characteristics of statistical instability

In a finite observation interval statistical instability of random sequence  $X_1, X_2, \dots, X_N$  or random process  $X(t)$ , both with limited first two moments, can be characterized by parameters describing fluctuation of average  $Y_N$  or fluctuation of expectation  $m_{y_N}$ .

Fluctuation of average  $Y_N$  is characterized by sample variance  $\bar{D}_{Y_N}$  of average. To characterize statistical instability of random sequence, the parameter of statistical instability  $\gamma_{1N} = \frac{M[\bar{D}_{Y_N}]}{D_{x_N}}$  may be used where  $M[\cdot]$

is expectation operator and  $D_{x_N} = \frac{1}{N} \sum_{n=1}^N D_{x_n}$  is mean of variances  $D_{x_n}$  of random variables  $X_n$ .

Fluctuation of expectation  $m_{y_N}$  is characterized by fluctuation variance  $\bar{D}_{m_{y_N}} = \frac{1}{N} \sum_{n=1}^N (m_{y_n} - \bar{m}_{m_{y_N}})^2$  of expectation of average  $Y_N$ , where  $\bar{m}_{m_{y_N}} = \frac{1}{N} \sum_{n=1}^N m_{y_n}$  is fluctuation average of expectation  $m_{y_n}$  of average  $Y_N$ .

Therefore another parameter of statistical instability may be the parameter  $\gamma_{2N} = \frac{\bar{D}_{m_{YN}}}{D_{x_N}}$  that is fluctuation variance of average expectation  $\bar{D}_{m_{YN}}$  normalized on mean  $D_{x_N}$  of variances of random variables  $X_n$ .

It follows from Chebyshev theorem, that parameter of statistical instability  $\gamma_{1N}$  tends to parameter of statistical instability  $\gamma_{2N}$  when  $N \rightarrow \infty$ . If the sequence is statistically stable, the value of these parameters tends to zero (because  $M[\bar{D}_{Y_N}] \rightarrow 0$  and  $D_{x_N}$  is a finite quantity) and if the sequence is statistically instable, the value of them tends to some positive number, fluctuates, or tends to infinite.

The values of statistical instability  $\gamma_{1N}$  and  $\gamma_{2N}$  depend from value of expectation of sample variance of average  $M[\bar{D}_{Y_N}]$ , variance of average expectation  $\bar{D}_{m_{YN}}$ , and mean  $D_{x_N}$  of variances of random variables  $X_n$ . With reducing of fluctuation parameters  $M[\bar{D}_{Y_N}]$  and  $\bar{D}_{m_{YN}}$ , parameters of statistical instability are reducing too, and with reducing of mean  $D_{x_N}$  of variances, parameters of statistical instability are rising.

Sometimes more useful are parameters of statistically instability  $\mu_{1N}$ ,  $\mu_{2N}$  linked with parameters  $\gamma_{1N}$  and  $\gamma_{2N}$  by the following expressions:  $\mu_{1N} = \sqrt{\frac{\gamma_{1N}}{1 + \gamma_{1N}}}$ ,  $\mu_{2N} = \sqrt{\frac{\gamma_{2N}}{1 + \gamma_{2N}}}$ . Less values of parameters  $\mu_{1N}$ ,  $\mu_{2N}$  more statistically stable is according sequence.

Parameters  $\gamma_{1N}$  and  $\gamma_{2N}$  have only lower bound (that is zero) and parameters  $\mu_{1N}$ ,  $\mu_{2N}$  have as lower as upper bounds: lower bound equals to zero and upper bound – to unit.

Calculating results of parameters  $\gamma_{1N}$  and  $\gamma_{2N}$  for described eight models and also four obtained records of line voltage are presented in Fig. 6. According calculating results of parameters  $\mu_{1N}$  and  $\mu_{2N}$  are viewed in Fig. 7.

Thin lines 1 – 2 and 7 – 8 present calculating results for models 1 – 2 and 7 – 8 accordingly, bold lines 3 – 6 present calculating results for models 3 – 6 accordingly, and heavy lines 9 – 12 present calculating results for obtained records of line voltage.

In calculations, sample variances were used instead of variances.

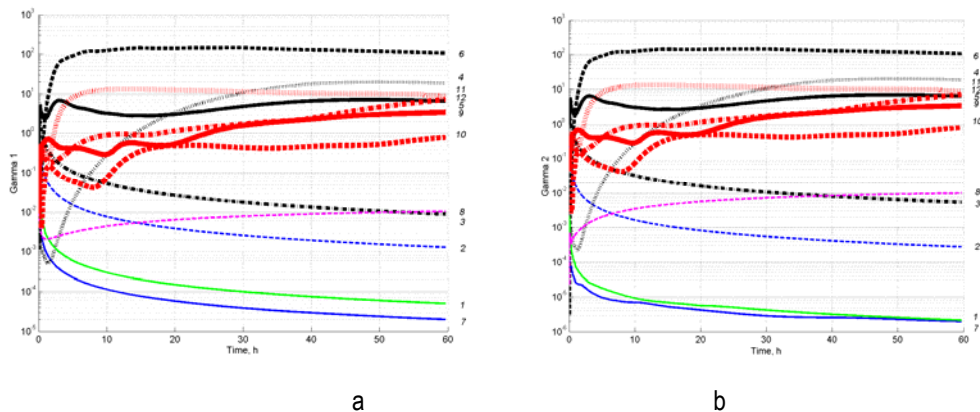


Fig. 6. Parameter of statistical instability of average  $\gamma_{1N}$  (a) and parameter of statistical instability of expectation of average  $\gamma_{2N}$  (b)

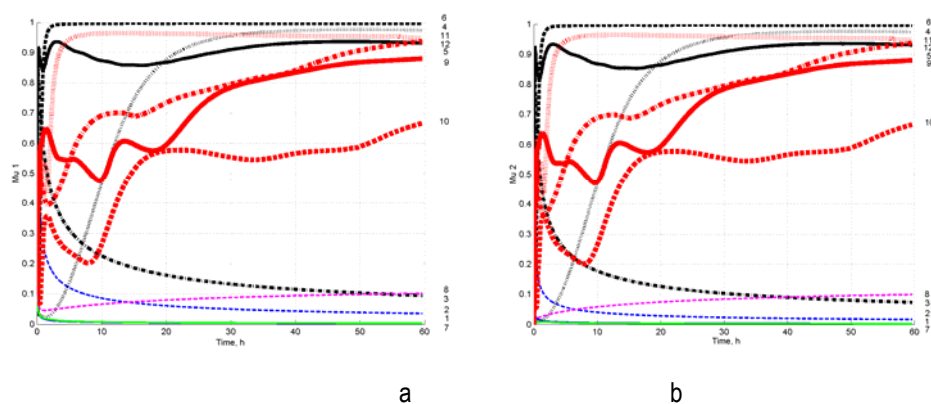


Fig. 7. Parameter of statistical instability of average  $\mu_{1N}$  (a) and parameter of statistical instability of expectation of average  $\mu_{2N}$  (b)

It follows from the curves, in the area of high observation time for all models and real processes, parameter of statistical instability of average  $\gamma_{1N}$  practically equals to parameter of statistical instability of expectation of average  $\gamma_{2N}$  ( $\gamma_{1N} \approx \gamma_{2N} = \gamma_N$ ) and parameter of statistical instability of average  $\mu_{1N}$  – to parameter of statistical instability of expectation of average  $\mu_{2N}$  ( $\mu_{1N} \approx \mu_{2N} = \mu_N$ ). So, any of these parameters may be used to describe statistical instability of the process.

Real processes can be presented as a mixture of statistical stable and statistical unstable components. Value of parameter  $\mu_N$  characterizes approximately percentage composition in the researched process of statistically unstable components. Time changes of value of this parameter give information about observation intervals on which the process can be regarded as statistically stable or statistically unstable.

For models 1 – 3 and 7, that accord to statistically stable processes, parameter value  $\mu_N$  steadily reduces with rising of observation time  $t$ ; for models 4 – 6 and 8, that accord to statistically unstable processes, value of this parameter increases. In the area of large observation times, for all real voltage processes the value of the parameter or steadily increases or oscillates on high level.

In the area of large observation times, values of parameters  $\gamma_{1N}$ ,  $\gamma_{2N}$  and  $\mu_{1N}$ ,  $\mu_{2N}$  are higher for statistically unstable processes than for statistically stable ones. This fact corroborates usefulness of parameters  $\gamma_{1N}$ ,  $\gamma_{2N}$  and  $\mu_{1N}$ ,  $\mu_{2N}$  for detection of statistical stability disturbance.

In the area of large observation times, values of instability parameters  $\gamma_{1N}$ ,  $\gamma_{2N}$  are in dozens times more for all experimental obtained records (not only presented in Fig. 6, 7) than for researched statistically stable models. It follows from this result that line voltage fluctuation has evidently statistically instable properties.

The interval on which parameters of statistical instability have high values, begins from some hours and follows to end of the records. So, the area of statistical instability is continues one, and covers range from some hours to not less 60 hours.

Stable character of disturbances of statistical stability of line voltage allows to assume that there are analogous disturbances of statistical stability in case of other physical (and may be not only physical) phenomena. Statistical stability of some others physical quantities are researched in the next sections.

## 6. Experimental research of statistical stability of height of sea waves and of their periods

Statistical stability of height of sea waves and of their periods was experimentally researched on the basis of statistical data obtained by Shirshov Oceanology Institute, RAS [ESIMO, 2010] from 2001 to 2003 years over a period of 15 months. Data was registered by wave station with interval from an hour to some hours.

Statistical instability parameters  $\mu_N$  were calculated for height of sea waves and for their periods (Fig. 8).

It follows from the figure that statistical instability parameters  $\mu_N$  have high value beginning from the zero reading, according the first ten observation hours. This means that in all observation interval, time dependences of height of sea waves and time dependences of periods of these waves have statistically unstable character. Therefore in time intervals, that are more ten hours statistical prediction of value of these parameters is practically impossible.

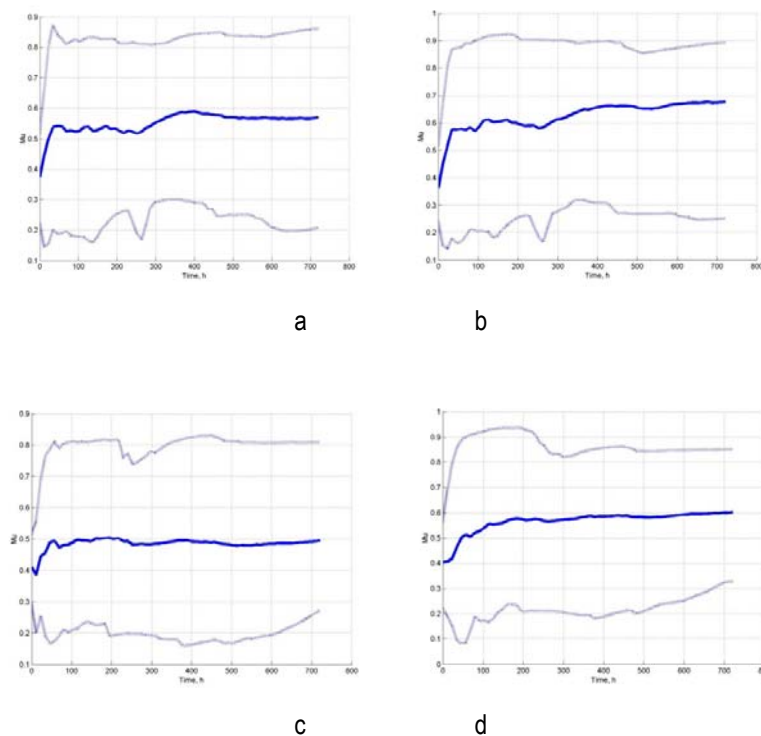


Fig. 8. Parameter of statistical instability  $\mu_N$  averaged on 15 months (solid lines) and bounds of the parameter (dotted lines): a – for maximum height of waves, b – for 10% height of waves, c – for periods of maximum height of waves, d – for periods of 10% height of waves

## 7. Experimental research of statistical stability of Earth magnetic field

Magnetic field of the Earth is changed in time and space. Observations of its fluctuation are led for a lot years in different Earth points.

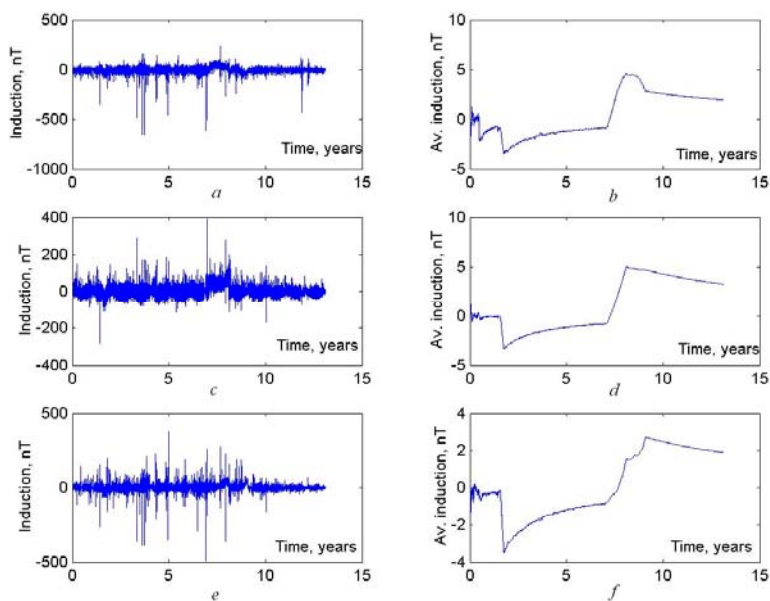


Fig. 9. Time dependencies of  $x$  ,  $y$  , and  $z$  components of Earth magnetic induction for 13 years in Moscow region (a, c, e) and according time dependencies of averages (b, d, f)

Time dependencies of  $x$  ,  $y$  , and  $z$  components of Earth magnetic induction obtained on the basis of IZMIRAN data [IZMIRAN, 2010] are presented in Fig. 9 a, c, e. According time dependencies of averages are shown in Fig. 9 b, d, f. Statistic instability parameters  $\mu_{1N}$  ,  $\mu_{2N}$  calculated on the described technique are presented in Fig. 10.

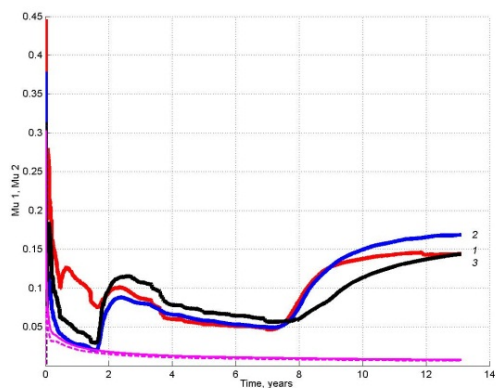


Fig. 10. Time dependencies of statistical unstable parameters  $\mu_{1N}$  ,  $\mu_{2N}$  for  $x$  ,  $y$  , and  $z$  components of Earth magnetic induction for 13 years in Moscow region (according 1, 2, and 3 curves) and also for control white Gaussian noise (curves without number). Solid lines accord to parameters  $\mu_{1N}$  and dotted lines – to parameters

$\mu_{2N}$



Analyze of curves shows that Earth magnetic induction field is not statistically stable, in general, although there are some time intervals where magnetic induction is practically stable. Duration of these intervals is not regular and fluctuates from some months to some years. Hence, statistical prediction of Earth magnetic induction field in the interval that is more some months is problematic and in the interval that is more some years is practically impossible.

## 8. Experimental research of statistical stability of currency rate

Imaging about statistical instability of currency rate gives curves in Fig. 11, obtained on the basis of FOREXITE data [FOREXITE, 2010]. It follows from curves, statistical instability parameter has high values beginning from the first observation hours and continuously rises. Hence, currency rate is statistically unstable quality and its statistical prediction is practically impossible.

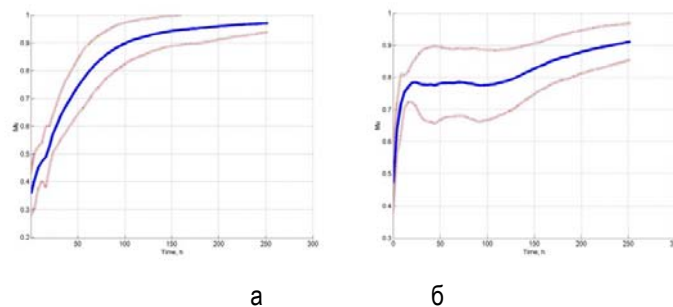


Fig. 11. Statistical instability parameter  $\mu_N$  averaged on 16 decades (solid line) and bounds of its changes determined by STD (dotted lines) for currency rate of AUD relative to USD for 2001 (a) and 2002 (b) years

## Conclusion

The materials generalizing results of author's researches in the field of disturbance of statistical stability have been presented. The concepts of statistical stability of random sequences and random processes are formalized. Parameters characterizing their statistical stability in finite observation interval are proposed. It is found that statistical unstable processes are especial class of non-stationary processes. Statistical stability disturbance can be explained by low frequency cyclical fluctuations of expectation. It has been researched statistical stability of a number of physical quantities and processes, in particular line voltage, height of sea waves and their periods, Earth magnetic field fluctuation, and currency rate. It has been found that for all researched quantities and processes, there is statistical stability disturbance on large observation intervals. Obtained results corroborate the hypotheses that are in base of new physico-mathematical theory of hyper-random phenomena: real physical events, quantities, processes, and fields are not statistical stable and they can be adequately described by hyper-random models taking into consideration their statistical instability.

## Bibliography

- [Alimov, Kravtsov, 1992] Yu.I.Alimov, Yu.A.Kravtsov. Is probability a "normal" physical quantity? // Physics-Uspekhi. – Vol. 5. – No 7, 1992. – P. 606 – 622.
- [Bernoulli, 1713] Jakob Bernoulli. The art of conjecturing. – 1713.
- [Crownover, 1995] R.M.Crownover. Introduction to fractals and chaos. Jones and Bartlett Pub., Inc., Boston – London. – 1995.
- [ESIMO, 2010] Единая государственная система информации об обстановке в мировом океане ЕСИМ. Данные Института океанологии им. П.П. Ширшова РАН. <http://ias.ocean.ru/esimo>.

- [Ferson, Kreinovich, Ginzburg, Myers, 2003] S.Ferson, V.Kreinovich, L.Ginzburg, D.S.Myers. Constructing probability boxes and Dempster-Shafer structures. SAND report SAND2002-4015. – 2003. – 143 p.
- [FOREXITE, 2010] FOREXITE. [www.forexite.com](http://www.forexite.com).
- [Gorban, 2008] I.I.Gorban. Hyper-random Phenomena: Definition and Description // International Journal "Information Theories and Applications". – No 15, 2008. – P. 203 – 211.
- [Gorban, 2009] I.I.Gorban. Cognition horizon and the theory of hyper-random phenomena. // International Journal "Information Theories and Applications". – 2009.
- [Hagan, Demuth, Beale, 1996] M.T.Hagan, H.B.Demuth, M.H. Beale. Neural network design. Boston. MA: PWS Publishing. – 1996.
- [ISO, 2006] International standard ISO 3534-1:2006(E/F). Statistics. Vocabulary and symbols. Part I: General statistical terms and terms used in probability. – 2006.
- [ZMIRAN, 2010] Данные о вариации индукции магнитного поля в районе Москвы. Институт земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН. <http://forecast.izmiran.rssi.ru/bankr.htm>.
- [Keller, 1986] J.B. Keller. The probability of heads // Am. Math. Monthly. Vol. 93, 1986. – P. 191.
- [Mises, 1964] R. von Mises. Mathematical theory of probability and statistics / Edited and complemented by H. Geiringer. – N.Y. and London: Acad. Press. – 1964. – 232 p.
- [Zadeh, Kacprzyk, 1992] L.A.Zadeh, J.Kacprzyk (Eds.). Fuzzy logic for the management of uncertainty. John Wiley & Sons. – New York. – 1992.
- [Алефельд, Херцбергер, 1987] Г.Алефельд, Ю.Херцбергер. Введение в интервальные вычисления. – М.: Мир. – 1987. – 356 с.
- [Гнеденко, 1961] Б.В.Гнеденко. Курс теории вероятностей. – М.: Изд-во физ.-мат. литературы. – 1961. – 406 с.
- [Горбань, 2007] И.И.Горбань. Теория гиперслучайных явлений. К.: ИПММС НАН Украины. – 2007. – 184 с. (эл. версия на сайте <http://ifsc.ualr.edu/jdberleant/intprob/>).
- [Горбань, 2010 (1)] И.И.Горбань. Нарушение статистической устойчивости физических процессов // Математические машины и системы. – №1, 2010. – С. 171 – 184.
- [Горбань, 2010 (2)] И.И.Горбань. Эффект статистической неустойчивости в гидрофизике. Труды десятой Всероссийской конференции «Прикладные технологии гидроакустики и гидрофизики» 25 – 27 мая 2010, Санкт-Петербург.
- [Горбань, 2010 (3)] И.И.Горбань. Статистическая неустойчивость магнитного поля Земли. Труды 6 дистанционной конференции «Системы поддержки принятия решений. Теория и практика. СППР '2010» 7 июня 2010 г., Киев.
- [Горбань, 2010 (4)] И.И.Горбань. Исследование статистической устойчивости курса валют. Труды пятой конференции «Математическое и имитационное моделирование систем» 22 – 26 июня 2010, Киев.
- [Колмогоров, 1936] А.Н.Колмогоров. Основные понятия теории вероятностей. – М.: ОНТИ. – 1936. – 175 с.
- [Колмогоров, 1986] А.Н.Колмогоров. О логических основаниях теории вероятностей // Теория вероятностей и математическая статистика. – М.: Наука. – 1986. С. 467 – 471.
- [Скорород, 1990] Предисловие А.В.Скоророда к монографии В.И.Иваненко, В.А.Лабковский. Проблема неопределенности в задачах принятия решения. – К.: Наукова думка. – 1990. – 134 с.
- [Тутубалин, 1972] В.Н.Тутубалин. Теория вероятностей. – М.: Из-во Московского университета. – 1972. – 230 с.
- [Шокин, 1981] Ю.И.Шокин. Интервальный анализ. – Новосибирск: Наука. – 1981. – 112 с.

---

## Authors' Information



**Igor Gorban** – Principal scientist of the Institute of Mathematical Machines and Systems Problem, National Academy of Sciences of Ukraine, Glushkov ave., 42, Kiev, 03187, Ukraine; e-mail: [igor.gorban@yahoo.com](mailto:igor.gorban@yahoo.com)

*Major Fields of Scientific Research: Theory of hyper-random phenomena, Space-time signal processing in complicated dynamic conditions, Fast multichannel space-time signal processing*

---

## MULTIPLE FOLDING OF VLSI REGULAR STRUCTURE VIA BOOLEAN SATISFIABILITY

Liudmila Cheremisinova

**Abstract:** *The problem under consideration is to reduce the area of the layout of regular VLSI structures by means of their folding. A novel reformulation of the folding problem as the Boolean satisfiability problem solved with any standard SAT-solver is proposed. The method is developed that allows creating a Boolean equation presenting folding problem in CNF form.*

**Keywords:** *design automation, area optimization, VLSI structure folding.*

**ACM Classification Keywords:** *B.6.3 [Logic Design]: Design Aids – Optimization; B.7.2 [Integrated circuits]: Design Aids –Layout.*

---

### Introduction

When designing control logic of custom VLSI, regular structured logic is widely used. This logic refers to logic forms that exhibit a high degree of regularity in their layout and interconnections. The widespread used hardware forms for structured design of digital VLSI systems are Programmable Logic Arrays (PLAs), Gate Matrices and, regular structures on the base of MOS-transistors connected in sequence (RMOS-structures) and others [Ullman, 1984]. All these forms have a two-dimensional structure consisting of an array of rows and columns. There are transistors in intersections of some rows and columns. The other special features of the structures under consideration are specified by the technology of VLSI manufacturing. The price paid for the structural regularity is larger chip area because obtained layouts are as a rule sparse: a large percentage of the row-column intersections are not personalized. Several techniques have been proposed for reducing the area required. Two approaches are usually used to reduce the area occupied by array-based structures: logic minimization that provides logic expressions with minimal number of products (and literals) and topological minimization reclaiming unused space.

The proposed paper deals with the problem of optimizing the area of the regular structure layout of regular structure by means of its folding [Hachtel, 1982], [DeMicheli, 1983], [Lecky, 1989], [Cheremisinova, 2004]. Folding is based on merging several columns (and/or rows) of an array-based structure into a single column (row). One special case of folding is simple column (and/or row) folding that involves merging a pair of columns (and/or rows) into a single column (row). The objective of one-dimensional column (row) folding is to find a permutation of the rows (columns) such that the set of columns (rows) could be placed in the minimum number of vertical lines. A generalization of the simple folding is multiple folding [DeMicheli, 1983] when more than two columns may share the same vertical line. This availability supposes rearrangement of different pieces (initial columns) of the same vertical line and thus multiple column (row) folding aims at determining a permutation of the regular structure rows (columns) which allow to implement in each vertical line (horizontal line) of the folded structure a set of initial columns (rows).

The proposed algorithm will be formulated regarding multiple column folding but it is valid for row folding too. Folding a column supposes to split that column into several segments so that several array inputs or outputs may share the same column of the folded two-dimensional structure. A multiple column folding is implemented by reordering the regular structure rows such that all of the rows populated along one column cover physically one segment of a vertical line that does not intersect with segments for other columns of the folding set.

Recent advances in solving Boolean satisfiability problems caused a significant resurgence of the application of satisfiability solvers (SAT-solvers) in different electronic design automation domains. In the last years, great improvements were achieved in both the speed and capacity of SAT-solvers [Eén], [Mahajan, 2005], [Goldberg, 2002], which are now very fast and can handle huge problems. The new efficient SAT-solvers open new possibilities for applying this technology by translating hard design problems to equivalent SAT problems. So

the existence of effective SAT-solvers makes it attractive to translate folding problem into Boolean problem solvable by SAT-solvers.

Devadas has been reformulated a routing problem as a Boolean satisfiability problem in his pioneer paper [Devadas, 1989], later the same manner the transformation of the problem of PLA column folding has been done in [Quintana, 1995]. Though the Boolean satisfiability is an NP-hard problem (as well as the folding problem [Luby, 1982]), there are successful, practical attacks on even large satisfiability problems by using Binary Decision Diagrams (BDDs) [Bryant, 1986] and heuristic search [Eén], [Mahajan, 2005], [Goldberg, 2002]. The papers [Devadas, 1989], [Quintana, 1995] attacked their problems using BDD representation of verified Boolean function whose internal structure captures the solved problem. In contrast to these approaches the method presented here turn a folding problem into an instance of a Boolean satisfiability problem solved with any standard SAT-solver. SAT-solvers normally operate on Boolean formulas in Conjunctive Normal Form (CNF), so the method is proposed that allows creating a Boolean function (that presents folding problem) in CNF form.

Boolean SAT formulations are binary in essence. Introduced Boolean variables represent solution alternatives, and Boolean formulas represent constraints imposed by the solved problem. All variable assignments satisfying Boolean formulas are equivalent when solving the satisfiability problem. During the search of SAT solution there is no cost mechanisms to favor one over another. Thus we can only get the answer whether some solution of our problem exists. That is why in [Quintana, 1995] the problem of optimum PLA folding is solved regarding a priori assigned folding size (number of folding pairs or sets). The problem of PLA folding is considered and formulated as a problem of Boolean satisfiability, deriving a Boolean function such that an assignment of variables that satisfy it (if it exists) defines a fold of the given size. Thus, they are forced continuously to reformulate the folding problem with decreasing values of folding sizes until a satisfiable problem formulation exists. Such a reformulation of the folding problem based on enumeration of folding size values becomes cumbersome for PLAs of great size, especially for “dense” PLAs.

Unlike that, in the present paper it is suggested to deal with the key problem of array-based structure folding – the problem of ascertaining whether the given folding set (the collections of columns) is implementable. There exist efficient methods of finding the folding sets (for example, in [Cheremisinova, 1999] this task is reduced to clique identification in graph of pairwise compatibility between columns) but the methods of examining the implementability of a collection of folding sets (they are observed in [Lecky, 1989]) are complicated and they are formulated concerning ordered folding collections only. Hence, in the paper we focus upon the task of reformulation the folding problem as SAT problem and solve it in more general statement. A collection of unordered column folding lists is dealt with, and the problem is reformulated as follows: it is necessary 1) to establish whether there exists ordering of elements within folding lists such that the collection of the obtained ordered column folding lists is implementable, and 2) to find the ordering of rows of the array-based structure induced by the collection of ordered column folding lists. The results described in the paper are generalization of ones proposed in [Cheremisinova, 2010] where the case of simple folding is considered and so the case of folding pairs is dealt with.

---

### **The problem of a regular structure folding**

---

The combinational PLA is an example of an array-based structure. The overall combinational PLA is a standard two level NOR-NOR structure. Its vertical lines are assigned with the input variables (and their complements) and output variables. Inputs run vertically through the first part of a PLA matrix, called as the PLA AND plane. It generates signals on its rows, which are used as inputs to the second part of a PLA matrix called as the PLA OR plane. In Figure 1 an example of an array-based structure (that can be viewed as PLA AND plane) and its folded form are shown. Here, a dot means placing a transistor on crosspoint of vertical and horizontal lines.

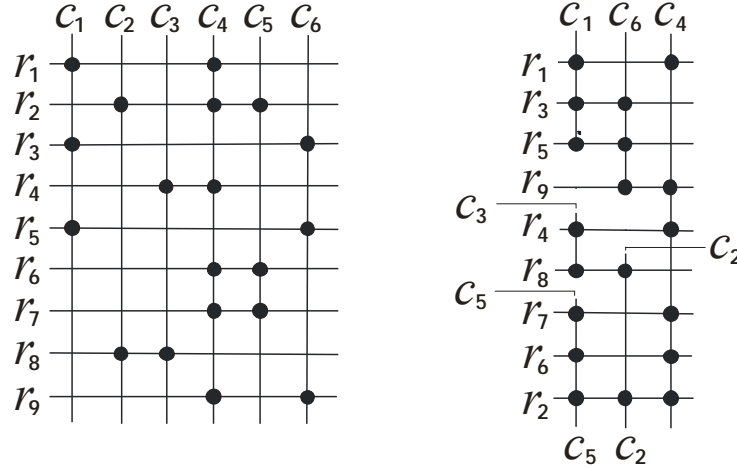


Figure 1 An example of array-based structure

Before we can give a mathematically tractable definition of PLA folding problem and its solution let give some other definitions. Either AND plane or both AND and OR planes together can be described in symbolic form by a Boolean matrix  $\mathbf{B}$  having sets  $C(\mathbf{B})$  and  $R(\mathbf{B})$  of its columns and rows. 1 in the position  $ij$  of the matrix  $\mathbf{B}$  means there is an appropriate crosspoint (transistor) between the  $i$ -row and the  $j$ -column in the matrix. Each column  $c_i$  of matrix  $\mathbf{B}$  implies the set  $R(c_i)$  of rows with a nonzero entry in this column:  $r_i \in R(c_i) \leftrightarrow b_{ij} = 1$ . And let  $C(r_i)$  be a set of columns with a nonzero entry in the  $i$ -th row of  $\mathbf{B}$ .

Any two columns  $c_i$  and  $c_j$  are *disjoint* if  $R(c_i) \cap R(c_j) = \emptyset$ . Two disjoint columns both do not have transistors on any particular row of the array-based structure. If there are no restrictions on the folding type two disjoint columns are compatible and can be folded together, so they form a column *folding pair*.

A column folding list  $I^c_k = (c_{k1}, c_{k2}, \dots, c_{km})$  is a set of disjoint (in pairs) columns  $c_{ij}$ , it is unordered in general case. An ordered column folding list  $I^{co}_k = \langle c_{k1}, c_{k2}, \dots, c_{km} \rangle$  is a column folding list  $I^c_k$  whose elements  $c_{ki}$  are ordered. An ordered column folding list (OCFL)  $I^{co}_k$  of cardinality two is an ordered column folding pair. Any OCFL  $I^{co}_k = \langle c_{k1}, c_{k2}, \dots, c_{km} \rangle$  can be actually implemented in the same vertical line of array-based structure moving  $c_{k1}$  above  $c_{k2}$ ,  $c_{k2}$  above  $c_{k3}$ , and so on,  $c_{k,m-1}$  above  $c_{km}$ . So OCFL  $I^{co}_k$  results to permutation on the set of rows:  $R(c_{k1}) > R(c_{k2})$  – the rows of  $R(c_{k1})$  are all above those of  $R(c_{k2})$ ,  $R(c_{k2}) > R(c_{k3})$  – the rows of  $R(c_{k2})$  are all above those of  $R(c_{k3})$ , and so on,  $R(c_{k,m-1}) > R(c_{km})$  – the rows of  $R(c_{k,m-1})$  are all above those of  $R(c_{km})$  inducing the relation on row set  $R(\mathbf{B})$ :

$$Pr(I^{co}_k) = \bigcup_{i,j} (R(c_{ki}) \times R(c_{kj})), \quad \text{i.e.} \quad Pr(I^{co}_k) = \{r_p \times r_q / r_p \in R(c_{ki}), r_q \in R(c_{kj}), i < j\}.$$

This relation is partial because it is irreflexive, asymmetric and transitive by its definition. Any unordered column folding list (CFL)  $I^c_k = (c_{k1}, c_{k2}, \dots, c_{km})$  induces up to  $m!$  possible OCFLs  $I^{co}_{ki} = \langle c_{k,i1}, c_{k,i2}, \dots, c_{k,im} \rangle$  begot by different permutations of columns in  $I^c_k$ . Thus CFP  $I^c_k$  induces one of  $m!$  possible partial relations on the row set  $R(\mathbf{B})$ . For example, the mentioned array-based structure (Figure 1) allows the CFL  $I^c = (c_1, c_3, c_5)$  that induces six OCFLs:

$$I^{co}_1 = \langle c_1, c_3, c_5 \rangle; I^{co}_2 = \langle c_1, c_5, c_3 \rangle; I^{co}_3 = \langle c_3, c_1, c_5 \rangle; I^{co}_4 = \langle c_3, c_5, c_1 \rangle; I^{co}_5 = \langle c_5, c_1, c_3 \rangle; I^{co}_6 = \langle c_5, c_3, c_1 \rangle. \quad (1)$$

and correspondingly six partial row orderings, the first of them is as follows:

$$\{r_1, r_3, r_5\} > \{r_4, r_8\}; \{r_4, r_8\} > \{r_2, r_6, r_7\}.$$

A column folding set (CFS)  $L^c_k = \{I^c_{k1}, I^c_{k2}, \dots, I^c_{kn}\}$  is a set of disjoint column folding lists (ordered  $L^{co}_k$  or unordered  $L^c_k$ ). The number of columns entering into all CFLs  $I^c_{ki} \in L^c_k$  is called as the size of CFS  $L^c_k$ . An ordered column folding set (OCFS)  $L^{co}_k = \{I^{co}_{k1}, I^{co}_{k2}, \dots, I^{co}_{kn}\}$  induces a set of ordering relations among the rows that is the union of ordering relations induced by OCFLs  $I^{co}_{ki}$  belonging to the OCFS  $L^{co}_k$ :

$$P^r(L^{\infty_k}) = \bigcup_{i=1}^n (P^r(I^{\infty_{ki}})).$$

This relation  $P^r(L^{\infty_k})$  is irreflexive, asymmetric but not transitive in general case. The transitive closure  $R^t(P^r)$  of  $P^r(L^{\infty_k})$  is irreflexive, transitive but can be not asymmetric.

It is proven [DeMicheli, 1983] that an OCFS  $L^{\infty_k}$  is *implementable* topologically (by a folded array-based structure) if the transitive closure  $R^t(P^r)$  of the relation  $P^r(L^{\infty_k})$  is a partial ordering on  $R(\mathbf{B})$ , that is  $R^t(P^r)$  is asymmetric. In other words, an OCFS  $L^{\infty_k}$  is implementable topologically if there exist linear order of the rows  $R(\mathbf{B})$  extending the row ordering  $P^r(L^{\infty_k})$ . For example, for the considered array-based structure (Figure 1) there are four CFLs:

$$I^c_1 = (c_1, c_3, c_5); I^c_2 = (c_1, c_2); I^c_3 = (c_2, c_6); I^c_4 = (c_3, c_6)$$

two of them are disjoint ( $I^c_1$  and  $I^c_3$ ) and may constitute CFS of maximal size:  $L^c = \{(c_1, c_3, c_5), (c_2, c_6)\}$ , it induces 12 OCFS  $L^{\infty_k}$  ( $I^c_1$  allows 6 different orderings (1) and  $I^c_2$  allows 2 different orderings:  $\langle c_2, c_6 \rangle$  and  $\langle c_6, c_2 \rangle$ ). One of the induced OCFS  $L^{\infty_1} = \{\langle c_1, c_3, c_5 \rangle, \langle c_2, c_6 \rangle\}$  generates the relation

$$\begin{aligned} P_1^r(L^{\infty_1}) &= R(c_1) \times R(c_3) \cup R(c_3) \times R(c_5) \cup R(c_2) \times R(c_6) = \\ &= \{r_1, r_3, r_5\} \times \{r_4, r_8\} \cup \{r_4, r_8\} \times \{r_2, r_6, r_7\} \cup \{r_2, r_8\} \times \{r_3, r_5, r_9\}. \end{aligned}$$

Its transitive closure  $R^t(P_1^r)$  contains conflict pairs  $(r_3, r_2)$  and  $(r_2, r_3)$ ; so  $R^t(P_1^r)$  is not a partial ordering on  $R(\mathbf{B})$  (because it is not asymmetric) and OCFS  $L^{\infty_1}$  is not implementable. But the other induced OCFS  $L^{\infty_2} = \{\langle c_1, c_3, c_5 \rangle, \langle c_6, c_2 \rangle\}$  generates the relation

$$\begin{aligned} P_2^r(L^{\infty_1}) &= R(c_1) \times R(c_3) \cup R(c_3) \times R(c_5) \cup R(c_6) \times R(c_2) = \\ &= \{r_1, r_3, r_5\} \times \{r_4, r_8\} \cup \{r_4, r_8\} \times \{r_2, r_6, r_7\} \cup \{r_3, r_5, r_9\} \times \{r_2, r_8\}, \end{aligned}$$

whose transitive closure  $R^t(P_2^r)$  is asymmetric, so it is partial relation and OCFS  $L^{\infty_2}$  is implementable.

An implementable OCFS  $L^{\infty_k}$  specifies the structure of the folded array, and its size is referred to as the size of

the folding: the number of OCFLs in  $L^{\infty_k}$  corresponds to the number of columns that will replace  $\sum_{i=1}^n |I^{\infty_{ki}}|$  columns of the initial array-based structure. For example, the size of implementable OCFS  $L^{\infty_2} = \{\langle c_1, c_3, c_5 \rangle, \langle c_6, c_2 \rangle\}$  is equal five and two columns of the folded regular structure replace five columns of the initial structure.

So the formal statement of optimal folding problem is as follows: given a Boolean matrix representing array-based structure, find an implementable ordered folding set of maximum size.

A column folding set (consisting of unordered folding lists)  $L^c_k = \{I^c_{k1}, I^c_{k2}, \dots, I^c_{kn}\}$  is *implementable* if there exists an implementable OCFS  $L^{\infty_k}$  which OCFLs  $I^{\infty_{ki}}$  are got by ordering CFLs  $I^c_{ki} \in L^c_k$ .

---

### OCFS implementability checking via Boolean satisfiability formulation

---

Further the problem is stated as follows: given a column folding set (ordered)  $L^{\infty} = \{I^{\infty_1}, I^{\infty_2}, \dots, I^{\infty_n}\}$ , it is necessary to verify whether it is implementable. If it is, the corresponding partial order relation on the regular structure rows will be found. OCFS  $L^{\infty}$  can be obtained by one of the known methods (for example, by the method from [DeMicheli, 1983] or from [Cheremisinova, 1999]) where these task is reduced to maximum cliques identification in the graph of compatibility relation between array columns).

As it follows from above discussion an OCFL  $I^{\infty_k} = \langle c_{k1}, c_{k2}, \dots, c_{km} \rangle$  induce the following partial ordering:

$$(R(c_{k1}) > R(c_{k2})) \cap (R(c_{k2}) > R(c_{k3})) \cap \dots \cap (R(c_{k,m-1}) > R(c_{km})) \quad (2)$$

The task of checking whether OCFS  $L^\infty$  is implementable is reduced as shown below to checking whether there exists such an ordering of rows that provides fulfillment of the condition (2) for all OCFL  $l^{co}_k \in L^\infty$ . Let us show how to reduce the task to solving logic equation.

Let form a set  $R_P$  of rows that enter into at least one of the sets  $R(c_{ij})$  ( $c_{ij} \in l^{co}_i$ ,  $l^{co}_i \in L^\infty$ ,  $i \in \{1, 2, \dots, n\}$ ,  $j \in \{1, 2, \dots, i_m\}$ ). Positions of rows of the set  $R(\mathbf{B}) \setminus R_P$  in the folded array-based structure have no influence on the possibility of the structure folding regarding column folding lists from  $L^\infty$ . Therefore we may consider further only rows of  $R_P$ . Moreover, we do not need to consider the rows each of which is a member of the only  $R(c_{kj})$  ( $c_{kj} \in l^{co}_k$ ,  $l^{co}_k \in L^\infty$ ), they may be excluded from  $R_P$  because their placing is fixed only relative to the rows of a single set  $R(c_{ki})$  and has no conflicts regarding placing rows from other sets  $R(c_{sq})$  ( $s \neq k$ ). In this way we can reduce the number of variables and terms in a formed equation.

To determine ordering of  $|R_P|$  rows we encode each of the rows of the set  $R_P$  in  $q = \lceil \log_2 |R_P| \rceil$  Boolean variables  $x_1, x_2, \dots, x_q$  (where  $\lceil t \rceil$  is the least integer that is not less than the value of  $t$ ). The value of the code  $x^1 x^2 \dots x^q$  for the row  $r_i$  specifies its serial number that shows in which physical horizontal line the row  $r_i$  will be in the folded array-based structure. Thus the conditions implied by orderings  $r_i > r_j$  or  $r_i < r_j$  can be represented by the following functions:

$$f_{ij}^> = (x^1_i x^2_i \dots x^q_i) > (x^1_j x^2_j \dots x^q_j) \quad \text{or} \quad f_{ij}^< = (x^1_i x^2_i \dots x^q_i) < (x^1_j x^2_j \dots x^q_j). \quad (3)$$

So the restriction (2), induced by an OCFL  $l^{co}_k = \langle c_{k1}, c_{k2}, \dots, c_{km} \rangle$  and imposed on the order of array rows, takes the following form:

$$\bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{kl})} \left( \bigcap_{r_j \in R(c_{kp})} f_{ij}^> \right) \right) \right). \quad (4)$$

An ordered folding set  $L^\infty = \{l^{co}_1, l^{co}_2, \dots, l^{co}_n\}$  is implementable if the following equation is satisfiable:

$$\bigcap_{k=1}^n \left( \bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{kl})} \left( \bigcap_{r_j \in R(c_{kp})} f_{ij}^> \right) \right) \right) \right) = 1 \quad (5)$$

If the equation (5) has an assignment to the variables  $x^i_j \rightarrow \{0, 1\}$  that makes it to equal 1 than the considered OCFS  $L^\infty$  is implementable. And the satisfying assignment specifies the serial numbers of the rows of  $R_P$ . Otherwise, when no satisfying assignment for the equation (5) exists, the considered OCFS  $L^\infty$  is not implementable.

For illustration the technique consider OCFS  $L^{co}_2 = \{\langle c_1, c_3, c_5 \rangle, \langle c_6, c_2 \rangle\}$  from the example above. As  $R(c_1) = \{r_1, r_3, r_5\}$ ;  $R(c_3) = \{r_4, r_8\}$ ;  $R(c_5) = \{r_2, r_6, r_7\}$ ;  $R(c_2) = \{r_2, r_8\}$ ;  $R(c_6) = \{r_3, r_5, r_9\}$  we have  $R_P = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$  that can be reduced to  $R_P = \{r_2, r_3, r_5, r_8\}$  because some rows enter into the only set  $R(c_i)$ . Thus we have  $R'(c_1) = \{r_3, r_5\}$ ;  $R'(c_3) = \{r_8\}$ ;  $R'(c_5) = \{r_2\}$ ;  $R'(c_2) = \{r_2, r_8\}$ ;  $R'(c_6) = \{r_3, r_5\}$ . Four rows from  $R_P$  are encoded using two Boolean variables  $x_2$  and  $x_1$ , their values for a row  $r_i$  are  $x^2_i$  and  $x^1_i$ . After encoding equation (5) takes the form:

$$\begin{aligned} & (f_{3,8}^> \wedge f_{5,8}^> \wedge f_{8,2}^>) \wedge (f_{3,2}^> \wedge f_{3,8}^> \wedge f_{5,2}^> \wedge f_{5,8}^>) = f_{3,8}^> \wedge f_{5,8}^> \wedge f_{8,2}^> \wedge f_{3,2}^> \wedge f_{5,2}^> = \\ & = (x^2_3 x^3_3 > x^2_5 x^3_5) \wedge (x^2_5 x^3_5 > x^2_8 x^3_8) \wedge (x^2_8 x^3_8 > x^2_2 x^3_2) \wedge (x^2_3 x^3_3 > x^2_2 x^3_2) \wedge (x^2_5 x^3_5 > x^2_2 x^3_2) \end{aligned} \quad (6)$$

### CFS implementability checking via Boolean satisfiability formulation

Further we consider that unordered column folding set  $L^c = \{l^c_1, l^c_2, \dots, l^c_n\}$  is given where  $l^c_k = (c_{k1}, c_{k2}, \dots, c_{km})$ . The problem is stated as follows: verify whether CFS  $L^c$  is implementable and, if it is, the appropriate implementable OCFS should be found and the corresponding partial order relation on the array rows. This case differs from that considered above only by what to check CFS implementability we should look over different orders of columns within all CFLs  $l^c_i \in L^c$ , i.e. we should generate all permutations on a set of columns for each folding list  $l^c_i \in L^c$ .

That follows from the fact that a CFL  $I^c_k = (c_{k1}, c_{k2}, \dots, c_{km})$  induces the set  $P(k_m)$  of size  $k_m!$  of possible OCFLs  $I^{co}_{ki} = (c_{k,i1}, c_{k,i2}, \dots, c_{k,im})$  ( $m = 1, 2, \dots, k_m!$ ) corresponding to different permutations of columns in  $I^c_k$ . Thus CFL  $I^c_k$  induces the following partial relations on the row set  $R(\mathbf{B})$  that is the extension of (2):

$$\bigcup_{I^{co}_{k,s} \in P(k_m)} ((R(c_{k,s1}) > R(c_{k,s2})) \cap (R(c_{k,s2}) > R(c_{k,s3})) \cap \dots \cap (R(c_{k,sm-1}) > R(c_{k,sm}))). \quad (7)$$

After encoding rows of the set  $R_P$  formed the same manner as in previous section the condition for CFL  $I^c_k$  takes the following form that is more complex than that of (4) for OCFL  $I^{co}_k$ :

$$\bigcup_{I^{co}_{k,s} \in P(k_m)} \left( \bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{k,sl})} \left( \bigcap_{r_j \in R(c_{k,sp})} f_{ij}^> \right) \right) \right) \right). \quad (8)$$

With (8) in mind, by analogy with (5), an unordered folding set  $L^c = \{I^c_1, I^c_2, \dots, I^c_n\}$  is implementable if the following equation is satisfiable:

$$\bigcap_{k=1}^n \left( \bigcup_{I^{co}_{k,s} \in P(k_m)} \left( \bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{k,sl})} \left( \bigcap_{r_j \in R(c_{k,sp})} f_{ij}^> \right) \right) \right) \right) \right) = 1 \quad (9)$$

If the equation (9) has an assignment to the variables  $x_{ij}^> \rightarrow \{0,1\}$  that makes it to equal 1 than the considered CFS  $L^c$  is implementable. And the satisfying assignment specifies the serial numbers of the rows of  $R_P$ , which show, in its turn, how the CFLs  $I^c_k = (c_{k1}, c_{k2}, \dots, c_{km})$  should be ordered. Otherwise, when no satisfying assignment for the equation (9) exists, the considered CFS  $L^c$  is not implementable.

To illustrate the technique let consider CFS  $L^c = \{(c_1, c_3, c_5), (c_2, c_6)\}$  from the above example. Thus we have  $R_P = \{r_2, r_3, r_5, r_8\}$ ,  $R'(c_1) = \{r_3, r_5\}$ ;  $R'(c_3) = \{r_8\}$ ;  $R'(c_5) = \{r_2\}$ ;  $R'(c_2) = \{r_2, r_8\}$ ;  $R'(c_6) = \{r_3, r_5\}$ , CFS  $L^c$  induces 12 OCFLs  $I^{co}_k$ . After encoding four rows from  $R_P$  the equation (9) takes the form:

$$\begin{aligned} & ((f_{3,8}^> \wedge f_{5,8}^> \wedge f_{8,2}^>) \vee (f_{3,2}^> \wedge f_{5,2}^> \wedge f_{2,8}^>) \vee (f_{8,3}^> \wedge f_{8,5}^> \wedge f_{3,2}^> \wedge f_{5,2}^>) \vee (f_{8,2}^> \wedge f_{2,3}^> \wedge f_{2,5}^>) \vee \\ & \vee (f_{2,3}^> \wedge f_{2,5}^> \wedge f_{3,8}^> \wedge f_{5,8}^>) \vee (f_{2,8}^> \wedge f_{8,3}^> \wedge f_{8,5}^>)) \wedge ((f_{2,3}^> \wedge f_{2,5}^> \wedge f_{8,3}^> \wedge f_{8,5}^>) \vee (f_{3,2}^> \wedge f_{3,8}^> \wedge f_{5,2}^> \wedge f_{5,8}^>)) \end{aligned} \quad (10)$$

### Transforming the equation for OCFS implementability checking into CNF form

When testing CFS whether it is implementable we need to ascertain SAT and to find a solution (if it exists), that gives OCFS and array row ordering. Effective SAT-solvers have been developed to perform the search for a satisfying solution (or prove that it does not exist). The best of SAT-solvers that exist today ([Eén], [Mahajan, 2005], [Goldberg, 2002] and others) are capable of handling millions of logical clauses and thousands of variables. SAT-solvers usually require their input to be in CNF form, thus it is necessary to convert the left part of the equations (5) and (9) into CNF form. CNF formula is expressed as the product of clauses, where each clause is the sum of literals. A literal is either variable or its negation.

To convert (5) into CNF form it is enough to get CNF form for the function  $f_{ij}^> = (x'_q x'_{q-1} \dots x'_1) > (x'_q x'_{q-1} \dots x'_1)$  (3) only. For example for  $q = 1$  (only one encoding variable) we have two-argument function  $f_{ij}^> = (x'_1) > (x'_1)$ , its minimum CNF formula is written over two variables and has two clauses:

$$C(f_{ij}^>) = x'_1 \bar{x}'_1.$$

For  $q = 2$  we have four-argument function  $f_{ij}^> = (x'_2 x'_1) > (x'_2 x'_1)$ , its minimum CNF formula is written over four variables and has five clauses:

$$C(f_{ij}^>) = (x'_2 \vee x'_1) \wedge (\bar{x}'_2 \vee x'_1) \wedge (x'_2 \vee \bar{x}'_1) \wedge (\bar{x}'_2 \vee \bar{x}'_1) \wedge (x'_2 \vee \bar{x}'_2).$$

In general case CNF formula for  $q = t$  depends on  $2t$  arguments. It is not difficult to show that the CNF formula for  $q = t + 1$  can be recursively received from the CNF formula for  $q = t$  by transforming each  $t$ -th clause  $d^t_r$  of the first CNF into two clauses  $(x'_{t+1} \vee d^t_r)$ ,  $(\bar{x}'_{t+1} \vee d^t_r)$  and adding into the resulting CNF one more additional clause



$(x'_{t+1} \vee \bar{x}'_{t+1})$ . So the minimum CNF for  $q = t + 1$  will consist of  $n_{t+1} = 2n_t + 1$  clauses, where  $n_t$  is the number of clauses in CNF for  $q = t$ .

After substituting CNFs for the function  $f_{ij}^>$  we transform (5) into the following Boolean equation whose left part is CNF:

$$\bigcap_{k=1}^n \left( \bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{kl})} \left( \bigcap_{r_j \in R(c_{kp})} C(f_{ij}^>) \right) \right) \right) \right) = 1 \quad (11)$$

Thus the task of checking whether an ordered column folding set is implementable is reduced to the task of checking whether CNF formula (11) is satisfiable.

To make clear the above transformations, it is useful to proceed with transformations of the formula (6) for OCFS  $L^{co_2} = \{<C_1, C_3, C_5>, <C_6, C_2>\}$  implementability checking into the following equation in CNF form:

$$\begin{aligned} f_{3,8}^> \wedge f_{5,8}^> \wedge f_{8,2}^> \wedge f_{3,2}^> \wedge f_{5,2}^> = & ((x^3_2 \vee x^3_1) \wedge (\bar{x}^8_2 \vee x^3_1) \wedge (x^3_2 \vee \bar{x}^8_1) \wedge (\bar{x}^8_2 \vee \bar{x}^8_1) \wedge (x^3_2 \vee \bar{x}^8_2)) \wedge \\ & \wedge ((x^5_2 \vee x^5_1) \wedge (\bar{x}^8_2 \vee x^5_1) \wedge (x^5_2 \vee \bar{x}^8_1) \wedge (\bar{x}^8_2 \vee \bar{x}^8_1) \wedge (x^5_2 \vee \bar{x}^8_2)) \wedge \\ & \wedge ((x^8_2 \vee x^8_1) \wedge (\bar{x}^2_2 \vee x^8_1) \wedge (x^8_2 \vee \bar{x}^2_1) \wedge (\bar{x}^2_2 \vee \bar{x}^2_1) \wedge (x^8_2 \vee \bar{x}^2_2)) \wedge \\ & \wedge ((x^3_2 \vee x^3_1) \wedge (\bar{x}^2_2 \vee x^3_1) \wedge (x^3_2 \vee \bar{x}^2_1) \wedge (\bar{x}^2_2 \vee \bar{x}^2_1) \wedge (x^3_2 \vee \bar{x}^2_2)) \wedge \\ & \wedge ((x^5_2 \vee x^5_1) \wedge (\bar{x}^2_2 \vee x^5_1) \wedge (x^5_2 \vee \bar{x}^2_1) \wedge (\bar{x}^2_2 \vee \bar{x}^2_1) \wedge (x^5_2 \vee \bar{x}^2_2)) = 1. \end{aligned} \quad (12)$$

Variable assignment satisfying the CNF is  $x^3_2 = x^3_1 = 1$ ;  $x^5_2 = 1$ ;  $x^5_1 = 0$ ;  $x^8_2 = 0$ ;  $x^8_1 = 0$ ;  $x^2_2 = x^2_1 = 0$ . So the rows of  $R_P = \{r_3, r_5, r_8, r_2\}$  are encoded as  $r_3 - 11$ ,  $r_5 - 10$ ,  $r_8 - 01$ ,  $r_2 - 00$ . The solution specifies partial row ordering  $r_3, r_5, r_8, r_2$  and linear ordering that is, for example,  $r_1, r_3, r_5, r_9, r_4, r_8, r_7, r_6, r_2$ , which leads to the folded array-based structure shown in Figure 1.

### Transforming the equation for CFS implementability checking into CNF form

For more general case of unordered column folding set, after substituting CNFs for the functions  $f_{ij}^>$  into (9) we construct following Boolean equation whose left part is not CNF generally speaking:

$$\bigcap_{k=1}^n \left( \bigcup_{l_{k,s}^{co} \in P(k_m)} C^>_{k,s} \right) = 1 \quad (13)$$

where  $C^>_{k,s}$  is the following CNF formula:

$$C^>_{k,s} = \bigcap_{l=1}^{m-1} \left( \bigcap_{p=l+1}^m \left( \bigcap_{r_i \in R(c_{k,sl})} \left( \bigcap_{r_j \in R(c_{k,sp})} C(f_{ij}^>) \right) \right) \right). \quad (14)$$

Boolean function on the left part of the equation (13) is almost CNF but not CNF in general case. To solve difficulty arising in transformation of Boolean formula  $(\bigcup_{l_{k,s}^{co} \in P(k_m)} C^>_{k,s})$  to a CNF form we suggest to convert it by

encoding  $k_m!$  CNFs  $C^>_{k,s}$  using unary codes, for example.  $k_m!$  additional coding Boolean variables  $Z_{k,s}$  are introduced for encoding  $k_m!$  CNFs  $C^>_{k,s}$ . To choose between  $k_m!$  CNFs  $C^>_{k,s}$  we can replace  $(\bigcup_{l_{k,s}^{co} \in P(k_m)} C^>_{k,s})$

(from (13) with:

$$\left( \bigcup_{l_{k,s}^{co} \in P(k_m)} (C^>_{k,s} \vee Z_{k,s}) \right) \wedge \left( \bigcup_{s=1}^{k_m!} \bar{Z}_{k,s} \right). \quad (15)$$

Here  $(C_{k,s}^> \vee Z_{k,s})$  can be considered as CNF that is simply deduced from  $C_{k,s}^>$  with the help of the distributive law (of Boolean algebra) of the conjunction concerning the disjunction: encoding variable  $Z_{k,s}$  is added into each clause  $d_m \in C_{k,s}^>$ , that is  $d_m \vee Z_{k,s}^>$ . To code all the CFLs  $l_k \in L^c$  we should introduce  $\sum_{k=1}^n (k_m!)$  encoding variables  $Z_{k,s}$ .

Thus, from the formulas (11), (13) – (15) obtained previously we construct CNF formula solving the key problem of array-based structure folding:

$$\bigcap_{k=1}^n \left( \bigcup_{l_{k,s}^c \in P(k_m)} (C_{k,s}^> \vee Z_{k,s}) \wedge \left( \bigcup_{s=1}^{k_m!} \bar{Z}_{k,s} \right) \right) = 1.$$

To make clear the above transformations, it is useful to proceed with transformations of the formula (10) for CFS  $L^c = \{(c_1, c_3, c_5), (c_2, c_6)\}$  implementability checking into the following equation in CNF form:

$$\begin{aligned} & ((f_{3,8}^> \wedge f_{5,8}^> \wedge f_{8,2}^>) \vee \dots \vee (f_{2,8}^> \wedge f_{8,3}^> \wedge f_{8,5}^>)) \wedge ((f_{2,3}^> \wedge f_{2,5}^> \wedge f_{8,3}^> \wedge f_{8,5}^>) \vee \dots) = \\ & = ((x^3_2 \vee x^3_1 \vee z^1_1) \wedge (\bar{x}^8_2 \vee x^3_1 \vee z^1_1) \wedge (x^3_2 \vee \bar{x}^8_1 \vee z^1_1) \wedge (\bar{x}^8_2 \vee \bar{x}^8_1 \vee z^1_1) \wedge (x^3_2 \vee \bar{x}^8_2 \vee z^1_1)) \wedge \\ & \wedge ((x^5_2 \vee x^5_1 \vee z^1_1) \wedge (\bar{x}^8_2 \vee x^5_1 \vee z^1_1) \wedge (x^5_2 \vee \bar{x}^8_1 \vee z^1_1) \wedge (\bar{x}^8_2 \vee \bar{x}^8_1 \vee z^1_1) \wedge (x^5_2 \vee \bar{x}^8_2 \vee z^1_1)) \wedge \\ & \wedge ((x^8_2 \vee x^8_1 \vee z^1_1) \wedge (\bar{x}^2_2 \vee x^8_1 \vee z^1_1) \wedge (x^8_2 \vee \bar{x}^2_1 \vee z^1_1) \wedge (\bar{x}^2_2 \vee \bar{x}^2_1 \vee z^1_1) \wedge (x^8_2 \vee \bar{x}^2_2 \vee z^1_1)) \wedge \dots \wedge \\ & \wedge ((x^2_2 \vee x^2_1 \vee z^6_1) \wedge (\bar{x}^8_2 \vee x^2_1 \vee z^6_1) \wedge (x^2_2 \vee \bar{x}^8_1 \vee z^6_1) \wedge (\bar{x}^8_2 \vee \bar{x}^8_1 \vee z^6_1) \wedge (x^2_2 \vee \bar{x}^8_2 \vee z^6_1)) \wedge \\ & \wedge ((x^8_2 \vee x^8_1 \vee z^6_1) \wedge (\bar{x}^3_2 \vee x^8_1 \vee z^6_1) \wedge (x^8_2 \vee \bar{x}^3_1 \vee z^6_1) \wedge (\bar{x}^3_2 \vee \bar{x}^3_1 \vee z^6_1) \wedge (x^8_2 \vee \bar{x}^3_2 \vee z^6_1)) \wedge \\ & \wedge ((x^8_2 \vee x^8_1 \vee z^6_1) \wedge (\bar{x}^5_2 \vee x^8_1 \vee z^6_1) \wedge (x^8_2 \vee \bar{x}^5_1 \vee z^6_1) \wedge (\bar{x}^5_2 \vee \bar{x}^5_1 \vee z^6_1) \wedge (x^8_2 \vee \bar{x}^5_2 \vee z^6_1)) \wedge \\ & \wedge (\bar{z}^1_1 \vee \bar{z}^2_1 \vee \bar{z}^3_1 \vee \bar{z}^4_1 \vee \bar{z}^5_1 \vee \bar{z}^6_1) \wedge \\ & \wedge ((x^2_2 \vee x^2_1 \vee z^1_2) \wedge (\bar{x}^3_2 \vee x^2_1 \vee z^1_2) \wedge (x^2_2 \vee \bar{x}^3_1 \vee z^1_2) \wedge (\bar{x}^3_2 \vee \bar{x}^3_1 \vee z^1_2) \wedge (x^2_2 \vee \bar{x}^3_2 \vee z^1_2)) \wedge \dots \wedge \\ & \wedge (\bar{z}^1_2 \vee \bar{z}^2_2) = 1. \end{aligned}$$

One of the variable assignments satisfying the CNF is the same as for CNF (12). It allows to encode rows of  $R_P = \{r_3, r_5, r_8, r_2\}$  as  $r_3 - 11, r_5 - 10, r_8 - 01, r_2 - 00$  and to fold array-based structure as shown in Figure 1.

---

## Conclusion

In this paper a novel reformulation of the multiple folding problem as the Boolean satisfiability problem solved with any SAT-solver was developed. To be exact the task of checking whether the given set of folding sets (or pairs in the case of simple folding) is implementable is considered. The distinctive features of the proposed method are as follows:

- 1) it can deal with a set of unordered column folding lists:
- 2) it orders column folding lists such a manner the resulting set of ordered folding lists to be implementable (if it is possible);
- 3) it specifies the ordering of rows of the regular structure induced by the found set of ordered column folding lists (if it exists).

---

## Acknowledgement

The research was supported by the Fond of Fundamental Researches of Belarus (Project  $\Phi 09K-025$ ).

---

## Bibliography

---

- [Ullman, 1984] Jeffrey D. Ullman Computational aspects of VLSI. Rockville, Md.: Computer Science Press, 1984, 495 p.
- [Hachtel, 1982] G.D. Hachtel, A.R. Newton and A.L. Sangiovanni-Vincentelli. An Algorithm for optimal PLA Folding. In: IEEE Trans. Computer-Aided Design of Integrated Circuit Syst., 1982, vol. CAD-1, No 2, pp. 63–77.
- [DeMicheli, 1983] G. DeMicheli and A. Sangiovanni-Vincentelli. A. Multiple Constrained Folding of Programmable Logic Arrays: Theory and Applications. In: IEEE Trans. Computer-Aided Design, 1983, vol. CAD-2, No 3, pp. 151–167.
- [Lecky, 1989] O.I. Lecky, O.I. Murphy, R.G. Absher. Graph theoretic algorithms for the PLA folding problem. In: IEEE Trans. Computer-Aided Design, 1989, vol. 8, No 9, pp. 1014–1021.
- [Cheremisinova, 2004] L. Cheremisinova. Simple folding of array-based VLSI structures. In: Proc. of 6-th Intern. Workshop on Boolean problems, Freiberg (Sachsen), Sept. 19–20, 2004, pp. 245–250.
- [Eén] N. Eén, and N. Sörensson. MiniSat. <http://www.cs.chalmers.se/Cs/Research/FormalMethods/MiniSat>.
- [Mahajan, 2005] Y. Mahajan, Z. Fu, and S. Malik. Zchaff2004: An Efficient SAT Solver. In: [Theory and Applications of Satisfiability Testing](#) (2004 SAT Solver Competition and QBF Solver Evaluation (Invited Papers)), Springer Berlin / Heidelberg, 2005, pp. 360–375.
- [Goldberg, 2002] E. Goldberg, and Y. Novikov. BerkMin: A Fast and Robust SAT-Solver. In: Design, Automation, and Test in Europe, March 2002, pp. 142–149.
- [Devadas, 1989] S. Devadas. Optimal Layout via Boolean Satisfiability. In Proc of International Conference on Computer-Aided Design (ICCAD '89), 1989, November, pp. 294–297.
- [Quintana, 1995] J.M. Quintana, M.J. Avedillo, M.P. Parra, and J.L. Huertas. Optimum PLA Folding through Boolean Satisfiability. In: Asian South Pacific Design Automation Conference (ASP\_DAC'95), 1995, pp. 289–293.
- [Luby, 1982] M. Luby, U. Vazirani, V. Vazirani, A. Sangiovanni-Vincentelli. Some theoretical results on the optimal PLA folding problem. In: Proc. of IEEE Conf on Circuits and Computers, 1982, pp. 165–170.
- [Bryant, 1986] R.E. Bryant. Graph-based algorithms for Boolean function manipulation. In: IEEE Trans. Computers, 1986, vol. C-35, No 8, pp. 677–691.
- [Cheremisinova, 1999] L.D. Cheremisinova. Area minimization of Weinberger matrix. In: Proc. of the Third Intern. Conf. on Computer-Aided Design of Discrete Devices, CAD DD'99, Minsk: Rep.of Belarus, Nov.10–12, 1999, vol. 1, pp. 80–87.
- [Cheremisinova, 1999] L.D. Cheremisinova. Some results in optimal PLA folding. In: Proc. of the Third Intern. Conf. on Computer-Aided Design of Discrete Devices, CAD DD'99, Minsk: Rep.of Belarus, Nov.10–12, 1999, vol. 1, pp. 59–64.
- [Cheremisinova, 2010] L.D. Cheremisinova. Regular structure folding based on solving logic equations. In: Proc. of 4<sup>th</sup> intern. Conf. of Tanaev's readings, March 28–29, 2010, Minsk: UIIP NAS of Belarus (in Russian: Л.Д. Черемисинова. Свертка регулярных структур на основе решения логических уравнений. В: Танаевские чтения. Доклады 4-й Междун.научной конф. (28–29 марта 2010 г., Мн.: ОИПИ НАН Беларуси).

---

## Authors' Information

---



**Liudmila Cheremisinova** – Principal Researcher, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: [cld@newman.bas-net.by](mailto:cld@newman.bas-net.by)

Major Fields of Scientific Research: Logic Design, CAD systems, optimization

## SIMPLE CONSTRAINED FOLDING OF PROGRAMMABLE LOGIC ARRAYS OF SPECIAL TYPE

Liudmila Cheremisinova, Irina Loginova

**Abstract:** *In the paper the constrained folding problem is investigated for programmable logic array (PLA) of special architecture type. Constraints applied to folding are given which are imposed by the used PLA structural features. A PLA simple folding algorithm is presented which takes into account the PLA structure peculiarities.*

**Keywords:** *design automation, area optimization, PLA structure folding.*

**ACM Classification Keywords:** *B.6.3 [Logic Design]: Design Aids – Optimization; B.7.2 [Integrated circuits]: Design Aids – Layout.*

**Conference topic:** *Mathematical methods research of complex systems.*

---

### Introduction

---

Structured logic refers to logic forms that exhibit a high degree of regularity in their layout and interconnections. One of widespread used regular structure is Programmable Logic Array (PLA) [Ullman, 1984]. The following are some of the reasons for wide use of PLA structures in chip design: they have regular structure that allows extensive use of automation in the design and routing and they are easy-to-modify. PLA is a popular, highly regular logic-structure and has a two-dimensional structure consisting of an array of rows and columns. There are transistors in intersections of some rows and columns. PLA layout design is easily automated because of its direct correspondence with PLA personality matrix. The price paid for the structural regularity is that much PLA area is unused because a large percentage of the row-column intersections are not personalized. Several techniques have been proposed for reducing the area required.

Existing PLA area minimization techniques can be divided into two categories: minimization of the functions being implemented without changing the input/output relationship, and restructuring the PLA matrix by reordering the rows and columns so that any empty space is eliminated. The first methods are concerned with two-level logic minimization of systems of Boolean functions [Brayton, 1984], [Zakrevskij, 2009] and the second ones with topological minimization of the PLA. Two-level logic minimization allows to reduce 1) the number of product terms and as the result the number of PLA rows, 2) the number of literals of product terms and as the result the number of transistors in PLA circuit, i.e. the PLA density.

Techniques of topological minimization reduce the number of physical columns and/or rows, they change the PLA structure by using a process called folding [Hachtel, 1982], [DeMicheli, 1983]. Folding is a technology-independent, it is developed for array structures and attempts to place two or more columns (and/or rows) together in the same physical vertical (and/or horizontal) line so that they can share this line. Column folding is said to be simple if utmost two columns (rows) share a single physical vertical (and/or horizontal) line. Otherwise, it is called multiple when more than two columns can share a single column.

Folding does not change the implemented logic in any manner, but reduces the number of columns (and/or rows), and thus reduces the area of the PLA. In the paper we deal with PLA folding and its effects on the PLA structure.

---

### The peculiarities of topological implementation PLA structures of used type

---

The structure of a PLA is congruent with the sum-of-products representation of Boolean functions. The overall combinational PLA consists of a group of rows and columns and is characterized by number of inputs, outputs and product terms. It is a standard two level NOR-NOR structure that can be divided into two planes called the AND-plane and the OR-plane. For every input signal, there are two rows (columns) carrying uncomplement and complement input literals in the AND-plane. The literal row of a signal is connected to its complement via an inverter, and so they are situated in the neighborhood. For every output signal, there is a row (column) in the OR-

plane. The columns running across both planes correspond to product terms. PLA AND-plane produces the product terms which are used by the OR-plane to form the sum-of products (SOPs) of implemented functions.

Figures 1 and 2 show the physical layouts of two different PLA types. The first one is the traditional PLA that does not allow folding, and the second one that allows simple row folding. The both topological PLA structures implement the same SOP system with 8 input, 8 output variables and 13 product terms (PLA misex1 from Benchmark'91 set). The core of the both PLAs consists of AND and OR matrices having approximately the same sizes. The PLA structures differ in sizes of input buffers that are situated on the right and on the left from the AND matrix. Thus input signals feed input buffers from the both PLA sides.

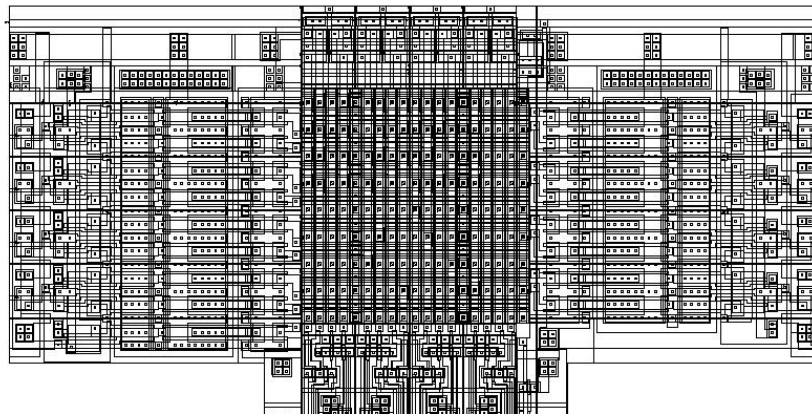


Figure 1. Topological layout of AND-plane of the traditional PLA

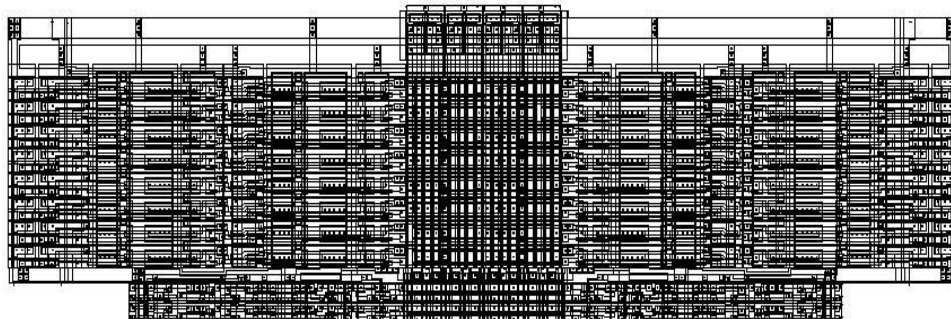


Figure 2. Topological layout of AND-plane of the foldable PLA

Each row of PLA AND-plane consists of two horizontal lines carrying signals corresponding to uncomplement and complement values of the same input variable. Each row of the traditional PLA (Figure 1) permits to be fed from the only PLA side. And a row of the foldable PLA (Figure 2) can take signals simultaneously from both PLA sides. In such a case these two signals share the same PLA row and such a manner we can realize PLA simple row folding. The same is true for the PLA OR-plane too, the matrix is not depicted in Figures 1 and 2 but it is organized in the same way.

From structural features of topological organization of foldable PLA structures some requirements follow. They must be taking into account when folding the PLA structure of foldable type.

1. Input signals corresponding to the literals  $x_i$  и  $\bar{x}_i$  of the same variable enter transistor pins from the same input buffer, i.e. they occupy two neighboring horizontal lines of the same AND-plane row. It is allowed to place row

carrying  $x_i$  above row carrying  $\bar{x}_i$  and vice versa, or to double some signals in the same PLA side or in the opposite side.

2. Each row of the PLA AND plane may be fed from both its sides when they can be folded each other. So two signals carrying  $x_i$  and  $\bar{x}_i$  arrive, for instance, from the right and other signals carrying  $x_j$  and  $\bar{x}_j$  arrive from the left. In this case the PLA line is divided into two segments separated from each other with the special break cell (in Figure 3 it is designated by a cross).

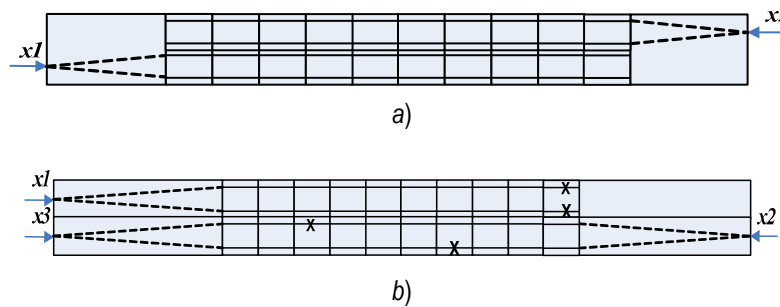


Figure 3. Sketch layout of the topology of rows of PLA AND-plane: a) for the case of traditional PLA; b) for the case of foldable PLA

3. When any variable does not participate in folding then it violates the other variable to feed from the opposite PLA side. To ensure blocking signal from the opposite PLA side the plug cell is introduced in this side.

4. In the case when one of the signals carrying  $x_i$  or  $\bar{x}_i$  may be fed and the other is blocked with plug cell (both lines of this row are used for inputs from the other PLA side and one of the literals  $x_j$  or  $\bar{x}_j$  is folded with  $x_i$  or  $\bar{x}_i$ ) then one more row should be added for the feeding one of  $x_i$  or  $\bar{x}_i$ .

5. PLA of the considered type permits only simple row folding, i.e. only two external signals may share the same PLA line.

### PLA-style logic structures and their folding

In the folding problem either AND-plane or both AND- and OR-planes together are described in symbolic form by a Boolean matrix  $\mathbf{B}$  having sets  $C(\mathbf{B})$  and  $R(\mathbf{B})$  of their columns (where uncomplement and complement modes of a variable have their own distinct column) and rows. 1 in the position  $i, j$  of the matrix  $\mathbf{B}$  means there is an appropriate crosspoint (transistor) between the  $i$ -th row and the  $j$ -th column in the matrix. Each row  $r_i \in R(\mathbf{B})$  implies a set  $C(r_i) \subseteq C(\mathbf{B})$  of rows, which are populated on it:

$$c_j \in C(r_i) \leftrightarrow b_{ij} = 1.$$

The objective of the matrix folding is to find the maximum number of pairs of rows that can be folded (each pair can be merged into a single line) simultaneously.

Any two rows  $r_i$  and  $r_j$  are disjoint if

$$C(r_i) \cap C(r_j) = \emptyset.$$

Two disjoint rows do not have transistors on the same particular PLA row. If there are no restrictions on the folding type two disjoint rows are compatible and can be folded together, so they form a row folding pair.

The objective of matrix folding is to arrange the rows and columns of the initial matrix in the plane using the least number of rows by means of assigning rows to the same horizontal lines.

Row folding can be obtained by permuting the columns of the PLA, so it introduces a restriction on the order of the columns. Any ordered row folding pair (ORF)  $\langle r_{k1}, r_{k2} \rangle$  can be actually implemented in the same horizontal PLA line moving  $r_{k1}$  to the left and  $r_{k2}$  to the right and resulting to permutation on the set of PLA columns: all

columns of  $C(r_{k1})$  are at the left of those of  $C(r_{k2})$ . So OFP  $p_k^o = \langle r_{k1}, r_{k2} \rangle$  induces a partial ordering  $C(p_k^o)$ :  $C(r_{k1}) < C(r_{k2})$  upon the columns of  $C(\mathbf{B})$ , that is specified by Cartesian product  $C(p_k^o) = C(r_{k1}) \times C(r_{k2})$ . Any unordered column folding pair (FP)  $(r_{k1}, r_{k2})$  corresponds to two possible OFPs  $\langle r_{k1}, r_{k2} \rangle$  or  $\langle r_{k2}, r_{k1} \rangle$  and thus induces one of two possible partial ordering relationships.

Let us have, for example, a simple PLA structure whose AND-plane is described by the following Boolean matrix:

	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$			
	1	1	0 0	0	0		$a$		
	0	0	1 0	0	1		$\bar{a}$		
	1	0	0 1	0	0		$b$		
<b>B</b>	=		$0$	$0$	$1$	$0$	$1$	$0$	$b\bar{}$
	0	0	0 1	0	0			$c$	
	1	0	1 0	1	0			$\bar{c}$	
	1	1	1 0	0	1			$d$	
	0	0	0 1	1	0			$\bar{d}$	

For this PLA there are eight unordered row folding pairs:

$(a, \bar{b}), (a, c), (a, \bar{d}), (\bar{a}, b), (\bar{a}, c), (\bar{a}, \bar{d}), (\bar{b}, c), (c, d)$

or 16 ordered row folding pairs. OFP  $p_{14}^o = \langle r_1, r_4 \rangle = \langle a, \bar{b} \rangle$  induces the following partial ordering

$R(p_{14}^o) = R(r_1) \times R(r_4) = \{(c_1, c_3), (c_1, c_5), (c_2, c_3), (c_2, c_5)\}$ .

A folding set (FS)  $P = \{p_1, p_2, \dots, p_s\}$  is a set of disjoint folding pairs (any row may enter only into one of pairs), and an ordered folding set (OFS)  $P^o = \{p^o_1, p^o_2, \dots, p^o_s\}$  is a set of disjoint ordered folding pairs. The relation  $C(P^o)$  induced by OFS  $P^o$  is defined as being the union of the relations induced by OFPs belonging to OFS  $P^o$ . But the restrictions imposed on the order of columns by one set of folded rows might conflict with the column ordering desired by another set of folded rows. Such conflicts on column orderings along with some other conditions (explained in the next section), make row folding an NP-complete optimization problem [Hachtel, 1982].

In [Hachtel, 1982] it is said that the ordered folding set  $P^o$  is implementable (has no conflicts on column orderings) if the transitive closure of  $\mathbf{C}_{P^o}^t$  of the ordering relation  $\mathbf{C}_{P^o}$  is a partial ordering relation on  $C(\mathbf{B})$ . Partially ordered set of columns that corresponds to  $\mathbf{C}_{P^o}$  can be embedded into a linearly ordered set of columns from  $C(\mathbf{B})$ . It is clear that implementable ordered folding set  $P^o$  contains all information needed to fold the matrix i.e. it specifies the pairs of rows to be folded together and their relative position (on the left or on the right). The objective is to find the implementable ordered folding set of maximum cardinality.

Thus folding requirements are as follows:

1) elements of a set of rows can be folded only if they are disjoint (there is no single column (product term) that uses both the rows (literals));

2) a set  $p_i$  of rows that can be folded should be compatible with a set  $P^o_i$  of rows that are already folded having in view that  $p_i$  should be compatible with respect to ordering of columns required by set  $P^o_i$ .

For example, the OFS  $\langle r_2, r_8 \rangle$  defines a partial ordering relation  $\mathbf{C}_{2,8}$  of the mentioned matrix  $\mathbf{B}$ :

$$\mathbf{C}_{2,8} = C(r_2) \times C(r_8) = \{(c_3, c_4), (c_3, c_5), (c_6, c_4), (c_6, c_5)\}.$$

For the matrix  $\mathbf{B}$  we can see that  $P^o = \{\langle r_2, r_8 \rangle, \langle r_5, r_7 \rangle\}$  implies

$$\mathbf{C}_{P^o}^t = \mathbf{C}_{2,8} \cup C(r_5) \times C(r_7) = \mathbf{C}_{2,8} \cup \{(c_4, c_1), (c_4, c_2), (c_4, c_3), (c_4, c_6)\}.$$

Its transitive closure  $\mathbf{C}_{P^o}^t$  is not a partial ordering relation on  $C(\mathbf{B})$  because there exist pairs  $(c_6, c_4)$  and  $(c_4, c_6)$ . Thus  $\{\langle r_2, r_8 \rangle, \langle r_5, r_7 \rangle\}$  is not implementable. In contradiction  $\{\langle r_2, r_8 \rangle, \langle r_7, r_5 \rangle\}$  is implementable.

### The peculiarities of folding of PLA structures of the considered type

As it was mentioned, the goal of simple row folding of PLA structures (without restrictions on the order of PLA columns and rows) is to find out an implementable ordered folding set of the greatest cardinality, i.e. having the most number of folding pairs. An ordered folding set  $P^o = \{po1, po2, \dots, po_s\}$  induces three-block partition  $(R1, R2, R3)$  of the set  $R(\mathbf{B})$  (and corresponding partition  $(X1, X2, X3)$  of the set of literals  $X$ ): the rows of  $R1$  and  $R2$  are the left and right segments of ordered folding pairs  $poi \in P^o$  and all other rows are included in  $R3$ .

The analysis of peculiarities of topological architecture of PLA structures of the considered type shows that the following variants of implementation of folding pairs from ordered folding set  $P$  are available.

1. Isolated realization. Any variable  $x_i$  (and  $\bar{x}_i$ ) occupies its own horizontal line (Figure 4, a) of the folded PLA if  $x_i \in X_3$  and  $\bar{x}_i \in X_3$ . For instance a signal comes to the AND-plane from the left and the break cell is situated on the right hand side of the line.
2. Coupled realization of two folding pairs. When  $\langle x_i^\sigma, x_j^\rho \rangle, \langle x_i^\sigma, x_j^\rho \rangle \in P^o$  ( $\sigma, \rho \in \{0, 1\}$  and  $x_i^1 = x_i, x_i^0 = \bar{x}_i$ ) then a distinct row (consisting of two horizontal lines) enters into AND-plane. It takes  $x_i$  and  $\bar{x}_i$  on the left but  $x_j$  and  $\bar{x}_j$  from the right (Figure 4, b). In this case compared with isolated realization (Figure 4, a), we have gain in PLA square equaled to one row or two lines.
3. Mixed realization. When  $\langle x_i^\sigma, x_j^\rho \rangle \in P^o$  but  $x_i^\sigma, x_j^\rho \in X_3$  then the pair  $\langle x_i^\sigma, x_j^\rho \rangle$  is subject to elimination from  $P^o$  (with carrying  $x_i^\sigma$  and  $x_j^\rho$  into  $X_3$ ) because the implementation of such a folding variant (Figure 4, c) does not give any gain in square in comparison with isolated realization of these two variables.
4. Combined realization. When  $\langle x_i^\sigma, x_j^\rho \rangle \in S$  and  $\langle x_i^0, x_j^0 \rangle \in P^o$  (or  $\langle x_j^0, x_i^0 \rangle \in P^o$ ,  $x_i^\sigma$  и  $x_i^0$  are allowed to be or not to be in  $X_3$ ) then PLA AND-plane has two rows fed by the same variable  $x_j$  (Figure 4, d). At the same time, the other two variables  $x_i$  and  $x_i$  take no additional PLA rows. In this case compared with isolated realization (Figure 4, a), we have gain in PLA square equaled to one row or two lines. But in comparison with coupled realization this gain should be divided on 3 (the number of participating variables) while in the case of the coupled realization it is divided on 2. As for combined realization, the variant when  $x_i^\sigma, x_i^0 \in X_3$  is more preferable as it allows to keep positive result of folding during topological implementation of folded PLA.

The purpose of row folding of PLA structure of considered type is to find out an implementable ordered folding set, not necessarily of the greatest cardinality, but providing the most gain in PLA square compared with not folded PLA. From the discussed above peculiarities of topological architecture of PLA structure it follows that the variant of coupled realization brings the most gain in comparison with the other variants, it provides saving 1 line per a variable. The second result is given by the combined realization, it provides saving 2/3 of line per a variable. The mixed realization as the variant of folding gives nothing compared with the isolated realization.



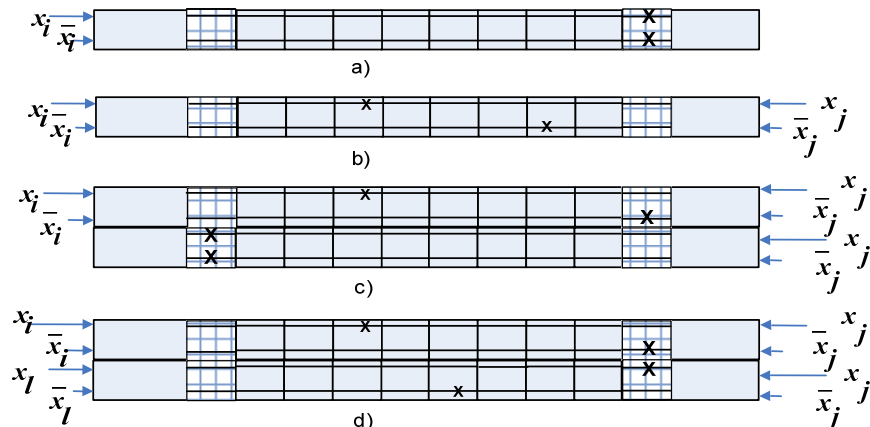


Figure 4. Sketch layouts of the variants of implementation of topology of PLA AND-plane rows

Having in view these reasons it is clear that the greatest reduction of PLA square can be achieved searching for implementable ordered folding set having as much as possible coupled folding pairs (variant 2) and, in the second turn, groups of combined three folding pairs (variant 4). The mixed realization (variant 3) gives nothing, so we may exclude from any ordered folding set each folding pair  $(r_{i1}, r_{i2})$  corresponding to  $(x_i^{\sigma}, x_j^{\rho})$  if  $x_i^{\sigma} \in X_3$  and  $x_j^{\rho} \in X_3$ . Such pairs are “redundant” when searching for folding set of the most cardinality for PLA of considered type.

### Method of simple row folding of PLA of considered type

A modification of the method [Cheremisinova, 1999] makes a good match for the PLA folding problem with discussed constraints on folding pairs. The method is based on the fact that implementable ordered folding set is a convex one: if an ordered folding set is implementable then any its subset is implementable too. The method supposes forming the compatibility relation on the set of folding pairs.

Taking into account that there is a great deal of folding pairs for rare PLA structures, here it is suggested to consider in initial stages of folding only unordered folding pairs. Any two unordered folding pairs  $(r_{i1}, r_{i2})$  and  $(r_{j1}, r_{j2})$  are compatible if unordered folding set (UFS)  $\{(r_{i1}, r_{i2}), (r_{j1}, r_{j2})\}$  is implementable. And UFS  $P = \{p_1, p_2, \dots, p_s\}$  is implementable if there exist an ordering of unordered folding pairs  $p_i \in P$  inducing implementable OFS  $P^o = \{p^{o_1}, p^{o_2}, \dots, p^{o_s}\}$ .

In [Cheremisinova, 1999] it was shown that

- 1) if UFS  $P$  is implementable it consists of pairwise compatible folding pairs;
- 2) unordered folding pairs  $(r_{i1}, r_{i2})$  and  $(r_{j1}, r_{j2})$  are not compatible iff the following conditions take place:

1.  $C(r_{i1}) \cap C(r_{j2}) \neq \emptyset$ ;
2.  $C(r_{j2}) \cap C(r_{i1}) \neq \emptyset$ ;
3.  $C(r_{i1}) \cap C(r_{j1}) \neq \emptyset$ ;
4.  $C(r_{j2}) \cap C(r_{j2}) \neq \emptyset$ .

(1)

If at least one of the conditions is violated the unordered folding pairs will be compatible.

The task of searching for PLA folding ensuring the most reduction of the area of PLA structure of considered type is broken into the following subgoals.

1. Searching for a set  $P^u$  of unordered folding pairs.
2. Excluding from  $P^u$  “redundant” folding pairs (as it is suggested in previous section).

3. Forming the compatibility relation  $\mathbf{Sp}$  on the set  $P^u$  of folding pairs.
4. Constructing a graph  $G = (V, E)$  of the compatibility relation  $\mathbf{Sp}$  where vertices from  $V$  correspond to the unordered folding pairs  $p_i \in P^u$  and two vertices  $v_i$  and  $v_j$  are connected with an edge  $e_{ij} \in E$  if corresponding pairs  $p_i$  and  $p_j$  are compatible.
5. Searching for maximal cliques  $G_i = (V_i, E_i)$  of the graph  $G = (V, E)$ .
6. Excluding from found cliques  $G_i = (V_i, E_i)$  vertices corresponding to "redundant" folding pairs (according to the folding pairs that are associated with the cliques).
7. Evaluating found maximal cliques  $G_i$  according to the criterion of saving of PLA lines (as in the previous section) and ordering the cliques  $G_i$  in accordance with these values.
8. Verifying whether unordered folding sets  $P_i^u$  corresponding to the cliques  $G_i$  are implementable and leaving out from further consideration cliques that violate this condition.
9. Choosing a clique  $G_i$  and the corresponding OFS of the most value of the criterion of saving PLA lines.

All the tasks except the tasks 5 and 8 have polynomial complexity, these two mentioned tasks are of exponential complexity.

For the example considered above we have eight unordered row folding pairs:  $P^u = \{(a, \bar{b}), (a, c), (a, \bar{d}), (\bar{a}, b), (\bar{a}, c), (\bar{a}, \bar{d}), (\bar{b}, c), (c, d)\}$ . So  $X_3 = \{c\}$  and there are no "redundant" folding pairs in  $P^u$ . Using (1) we define compatibility relation  $\mathbf{S}_p$  and then construct graph  $G = (V, E)$  of this relation shown in Figure 5.

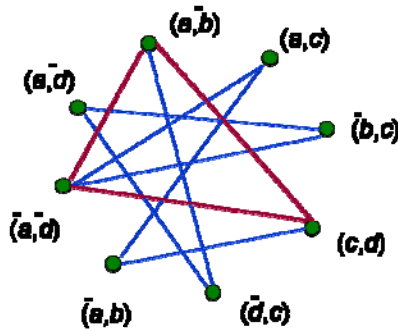


Figure 5. Graph  $G = (V, E)$  of the compatibility relation

There exists the only maximal clique in the graph  $G = (V, E)$  having three vertices:  $(a, \bar{b}), (\bar{a}, \bar{d}), (c, d)$ . For  $P^u = \{(a, \bar{b}), (\bar{a}, \bar{d}), (c, d)\}$  we get  $X_3 = \{b, \bar{c}\}$  and establish that there is no "redundant" folding pairs in this  $P^u$ . But as we have odd number of pairs, for the considered PLA type we can get gain only if we implement folding of two of them. So we should test on implementability two unordered folding sets:  $\{(a, \bar{b}), (\bar{a}, \bar{d})\}$  and  $\{(\bar{a}, \bar{d}), (c, d)\}$ , each of which has the same "binding" pair  $(\bar{a}, \bar{d})$ . Using the method [Cheremisina, 2010] of checking whether the given sets of unordered folding pairs are implementable we determine that they are implementable and get two following implementable sets of ordered folding pairs:

$$\{< \bar{a}, \bar{d} >, < d, c >\}, \{< a, \bar{b} > < \bar{d}, \bar{a} >\}.$$

The implementations of these two ordered folding sets are shown in Figure 6 in diagram form.

In Figure 7 more complex example of folding task is considered. PLA description *misex1* (Figure 7, a) from Benchmark'91 set is taken for folding. Its folded form is depicted in Figure 7, b where folding pairs are shown accompanied with the numbers of columns where break cells should be placed in the rows implementing folding pairs. Graphic form of implementation of the row folding is given in Figure 8.

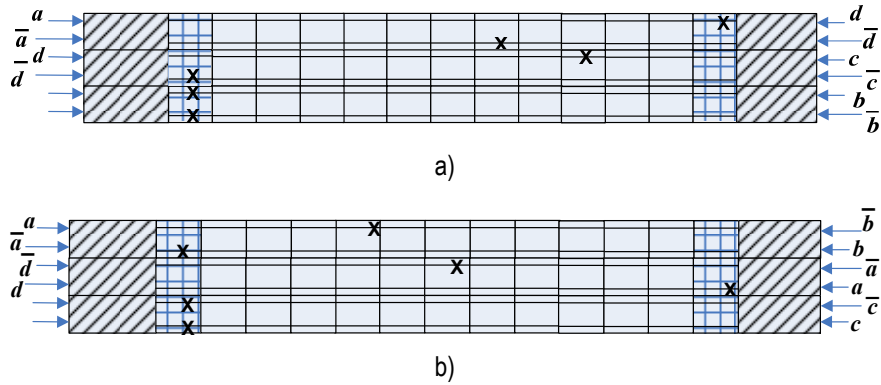


Figure 6. Sketch layouts of implementation of AND-plane columns from folding pairs:  
 a)  $\{ \langle a, d \rangle, \langle d, c \rangle \}$ ; b)  $\{ \langle a, b \rangle \langle d, a \rangle \}$

<pre>                 1             13  8  8 0111---- 10000100 1010---- 10111001 010----- 01011000 1001---- 01011100 0-00--1- 01011000 001--1-- 01011000 0011---- 01011000 010----0 00100000 01000--- 00100001 0010-0-- 00110100 01-1---- 00011000 0000--0- 00010010 0101--1 00000100     IN       a  ^a  b  ^b       c  ^c  d  ^d       e  ^e  f  ^f       g  ^g  h  ^h     OUT       y0  y1  y2  y3       y4  y5  y6  y7 ##### 00000000 #####             </pre>	<pre>             13 19     IN  11  5       1  b      0       2  ^b    g  9       3      ^g  0       4      ^f  0       5  d      f  7       6  ^d    ^h 11       7  a      h  4       8  ^a      0       9  ^c      0      10  c      ^e 9      11      e   0     OUT  8  0       1  y0     0       2  y1     0       3  y2     0       4  y3     0       5  y4     0       6  y5     0       7  y6     0       8  y7     0       P #####             </pre>
--	---

a)

b)

Figure 7. PLA misex1 folding: a) PLA misex1 description; b) description of the folded PLA

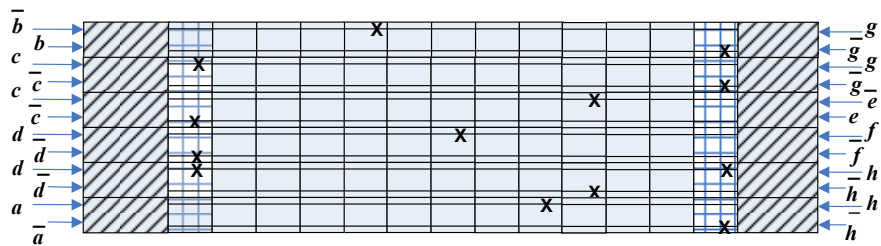


Figure 8. Sketch layout of implementation of topology of AND-plane rows of folded PLA misex1

---

**Conclusion**

---

The problem of reducing the area of the layout of PLA structures of a special type which allows row folding is investigated. The peculiarities of structural organization of this type PLA structures and its area optimization by means of folding are considered. Some restrictions imposed by the PLA topology on the folding mode are formulated. The task of PLA area optimization is formulated as constrained folding problem, and the method of simple row folding of PLA structures of considered type is proposed.

---

**Acknowledgement**

---

The research was supported by the Fond of Fundamental Researches of Belarus (Project  $\Phi 09K-025$ ).

---

**Bibliography**

---

- [Ullman, 1984] D. Jeffrey Ullman Computational aspects of VLSI. Rockville, Md.: Computer Science Press, 1984, 495 p.
- [Brayton, 1984] R.K. Brayton, G.D. Hachtel, C.T. McMullen et al. Logic minimization algorithm for VLSI synthesis. – Boston: Kluwer Academic Publishers, 1984, 193 p.
- [Zakrevskij, 2009] A. Zakrevskij, Yu. Pottosin, L. Cheremisinova. Optimization in Boolean space. – Tallinn: TUT PRESS, 2009, 241 p.
- [Hachtel, 1982] G.D. Hachtel, A.R. Newton and A.L. Sangiovanni-Vincentelli. An Algorithm for optimal PLA Folding. In: IEEE Trans. Computer-Aided Design of Integrated Circuit Syst., 1982, vol. CAD-1, No 2, pp. 63–77.
- [DeMicheli, 1983] G. DeMicheli and A. Sangiovanni-Vincentelli. A. Multiple Constrained Folding of Programmable Logic Arrays: Theory and Applications. In: IEEE Trans. Computer-Aided Design, 1983, Vol. CAD-2, No 3, pp. 151–167.
- [Cheremisinova, 1999] L.D. Cheremisinova. Some results in optimal PLA folding. In: Proc. of the Third Int. Conf. on Computer-Aided Design of Discrete Devices (CAD DD'99) (Minsk, Nov.10–12, 1999), 1999, Vol. 1, pp. 59–64.
- [Cheremisinova, 2010] L. Cheremisinova. Multiple Folding of VLSI Regular Structure via Boolean Satisfiability. In: this issue.

---

**Authors' Information**

---



**Liudmila Cheremisinova** – Principal Researcher, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: [cld@newman.bas-net.by](mailto:cld@newman.bas-net.by)

Major Fields of Scientific Research: Logic Design, CAD systems, optimization



**Irina Loginova** – Senior Researcher, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: [cld@newman.bas-net.by](mailto:cld@newman.bas-net.by)

Major Fields of Scientific Research: Topological Design, CAD systems

## APPLICATION OF PARETO OPTIMIZATION APPROACH TO OBSERVABILITY PROBLEM IN LINEAR AERODYNAMIC MODELS

**Andriy Zavorotnyy, Veda Kasyanyuk**

**Abstract:** Application of Pareto optimization based method of solving observability problem in aerodynamics is developed and considered. The error distribution of derivative approximation is investigated. The method takes into account the distribution of derivative approximation error. Systems with linear control are considered. The software for observation problem solving and research of considered method is developed and described. Results of method investigation with different data errors on generated data and on real flight data are given.

**Keywords:** Observability problem, Pareto optimization, aerodynamics, derivative approximation.

**ACM Classification Keywords:** I.6 Simulation and Modeling

---

### Introduction

Observability problem is the classical problem of control theory and consists in recovering the whole value of state-vector having known only some function of it:  $y = f(x)$ .  $f(x)$  is usually some linear combination of  $x$ . The problem can also be formulated as obtaining function  $G(y)$ , so that  $G(y) = G(f(x)) = E(x)$  is some estimate of  $x$  if such function  $G$  exists.

It is an urgent problem in aerodynamics, because all characteristics of aircraft cannot be measured by sensors during ordinary flight. While during test flights aircraft has more sensors, part of which is removed after testing.

A lot of methods, developed earlier are indeed not applicable because of their complexity and inadequacy in real needs.

The continuous model considered in this work is

$$\dot{x} = Ax + Bu \quad (1)$$

where  $x = (x_1(t), \dots, x_n(t))^T$  is the state-vector of object at  $t \in [0; T]$ , and  $u = (u_1(t), \dots, u_n(t))^T$  is the control-vector.  $A$  and  $B$  are real matrices of compatible dimensions.

The linear model is considered. Assume it's known some matrix  $C$  so that  $u = Cx$ , thus

$$\dot{x} = \bar{A}x \quad (2)$$

where  $\bar{A} = A + BC$ .

The classical observation problem for (2) is having at time  $\tilde{t} \in [0; T]$  defined first  $m$  elements of vector  $x$  to define it's unknown elements.

The Pareto optimization approach gives solution that satisfies several criteria representing different demands [Zavorotnyy, 2004], [Zavorotnyy, 2006].

As in this case it also considers errors in derivative estimation. That's why it was elected for solving observability problem.

---

### Notations

The derivative  $\dot{x}$  from (2) has to be estimated by values of  $x$ . Let's model the error in estimating  $\dot{x}$  considering

$$\dot{x}_{\tilde{t}} = \bar{A}x_{\tilde{t}} + \nu \quad (3)$$

where  $\dot{x}_{\tilde{t}}$  and  $x_{\tilde{t}}$  are vectors  $\dot{x}$  and  $x$  at  $\tilde{t}$ ,  $\nu = (\nu_1, \dots, \nu_n)^T$  is stochastic variable, characteristics of which are given below.

This vector can also be interpreted as sensor-measurement error in case of sensor-measured  $\dot{x}$ .  
Let the known part of  $\dot{x}_{\tilde{t}}$  from (2) is defined as

$$\tilde{x}_{\tilde{t}} = \tilde{A}x_{\tilde{t}} + \tilde{v} \quad (4)$$

where  $\tilde{x}_{\tilde{t}} = (\dot{x}_1(t), \dots, \dot{x}_m(t))^T$ ,  $\tilde{v} = (v_1(t), \dots, v_m(t))^T$ ,  $\tilde{A} = \begin{pmatrix} \bar{a}_{11} & \dots & \bar{a}_{1n} \\ \vdots & \ddots & \vdots \\ \bar{a}_{m1} & \dots & \bar{a}_{mn} \end{pmatrix}$ .

Then define

$$y = \tilde{x}_{\tilde{t}} - \tilde{A}_m x_{\tilde{t}}^m \quad (5)$$

where  $\tilde{A}_m = \begin{pmatrix} \tilde{a}_{11} & \dots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{m1} & \dots & \tilde{a}_{mn} \end{pmatrix}$  and  $\tilde{x}_{\tilde{t}} = (x_1(\tilde{t}), \dots, x_m(\tilde{t}))^T$ .

---

### Derivative approximation

As the method uses the covariance matrix of variable  $v$ , it's distribution has to be studied. In this work we consider system (1) to have oscillating solutions, which is typical case in aerodynamics problems. Thus the distribution is studied for basic trigonometric functions  $\sin(x)$  and  $\cos(x)$ . For both it appeared to have distribution

$$F_v(x) = \frac{1}{\pi} \arcsin\left(\frac{x}{\alpha}\right) \quad (6)$$

where  $\alpha$  is the absolute maximum of the error.

Consequently from (6),  $M(v) = 0$ , and

$$D(v) = \frac{\alpha^2 \pi}{2} \quad (7)$$

which can now be used in constructing covariance matrix after  $\alpha$  having been estimated.

Let covariance matrix is

$$\tilde{\mathfrak{R}}_v = \begin{pmatrix} D(v_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D(v_m) \end{pmatrix} \quad (8)$$

where values  $v_1 \dots v_m$  are supposed to be independent and  $D(v_i)$ ,  $i = \overline{1, m}$  is defined in (7).

---

### Pareto estimates

Substituting (4) in (5) we have :

$$y = \tilde{A}_{n-m} x_{\tilde{t}}^{n-m} + \tilde{v} \quad (9)$$

where  $\tilde{A}_{n-m} = \begin{pmatrix} \bar{a}_{1m+1} & \dots & \bar{a}_{1n} \\ \vdots & \ddots & \vdots \\ \bar{a}_{mm+1} & \dots & \bar{a}_{mn} \end{pmatrix}$  and  $\tilde{x}_{\tilde{t}}^{n-m} = (x_{m+1}(\tilde{t}), \dots, x_n(\tilde{t}))^T$ . Remind that for  $\tilde{v}$  expectation is 0 and it's

known covariance matrix (8) which is obviously reversible since (7).

In Pareto optimization approach given in [Zavorotnyy, 2004) the followed two-criteria problem applied to (9) is considered:

$$\begin{cases} h(G) = M \|G\tilde{v}\|^2 \rightarrow \min_G \\ \varphi(G) = \|G\tilde{A}_{n-m} - I\|^2 \rightarrow \min_G \end{cases} \quad (10)$$

where first criterion represents noise background of estimate, and second represents operator residual.  $G$  is a linear operator, so that  $G(y)$  would be the estimate of  $\tilde{x}_t^{n-m}$ .

According to considered method [Zavorotnyy, 2004], we obtain the following continuum of Pareto-optimal estimates from (10):

$$\hat{x}_t^{n-m} = \tilde{A}_{n-m}^T (\tilde{A}_{n-m} \tilde{A}_{n-m}^T + \mu \tilde{R}_v)^{-1} y \quad (11)$$

where  $\mu$  is the parameter of Pareto-optimization. The increase of  $\mu$  results in noise background decrease but also in operator residual [Zavorotnyy, 2004]. The criteria for defining value of  $\mu$  can be found in [Voronin, 1980].

## Software

The software implementing described approach (11) has been developed. It allows seeing the method in action both on real data, given in numeric representation as series of known parts of x-vector and on generated by given  $\bar{A}$  matrix data (3).

Here is given possibility to compare used method with two other classical methods of solving observability problem, which are also implemented in the software.

In the given examples errors were generated in normal distribution law.

Errors (variance = 0.01) in measurements of known parts of x-vector result in non-smoothness of estimated curve:

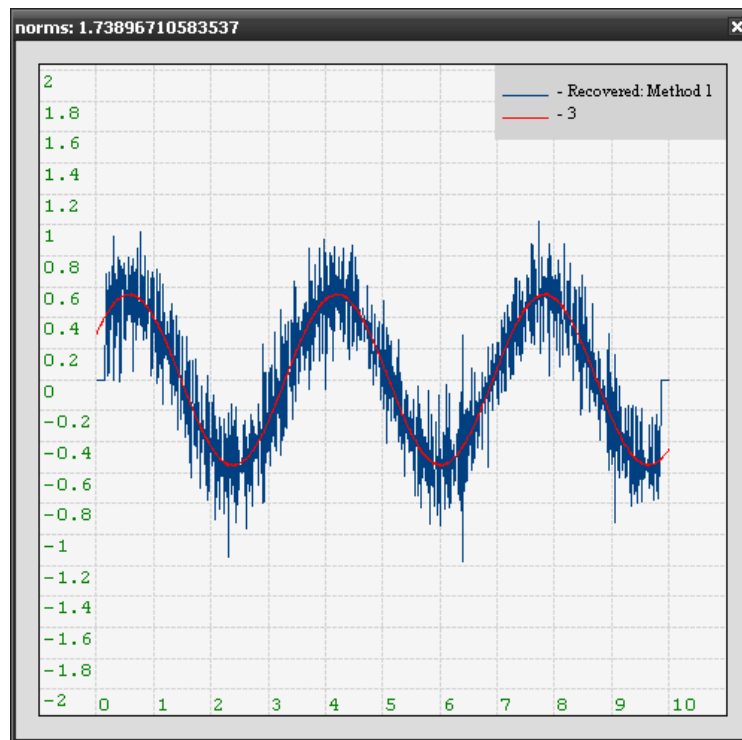


Figure 1: Non-smoothed estimation results.

The quality of curve is obviously much better after smoothing:

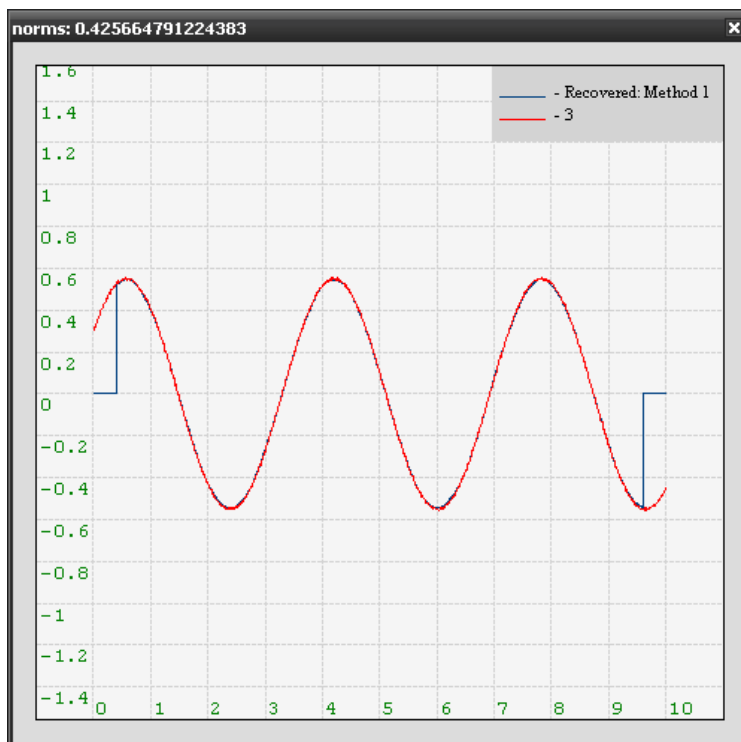


Figure 2: Smoothed estimation result.

Smoothness can substantially increase the quality of estimation because of zero expectation error of derivative estimation.

In the case of data measurement errors (variance = 0.0001), when classical methods are inconsistent, considered method gives perfect results:

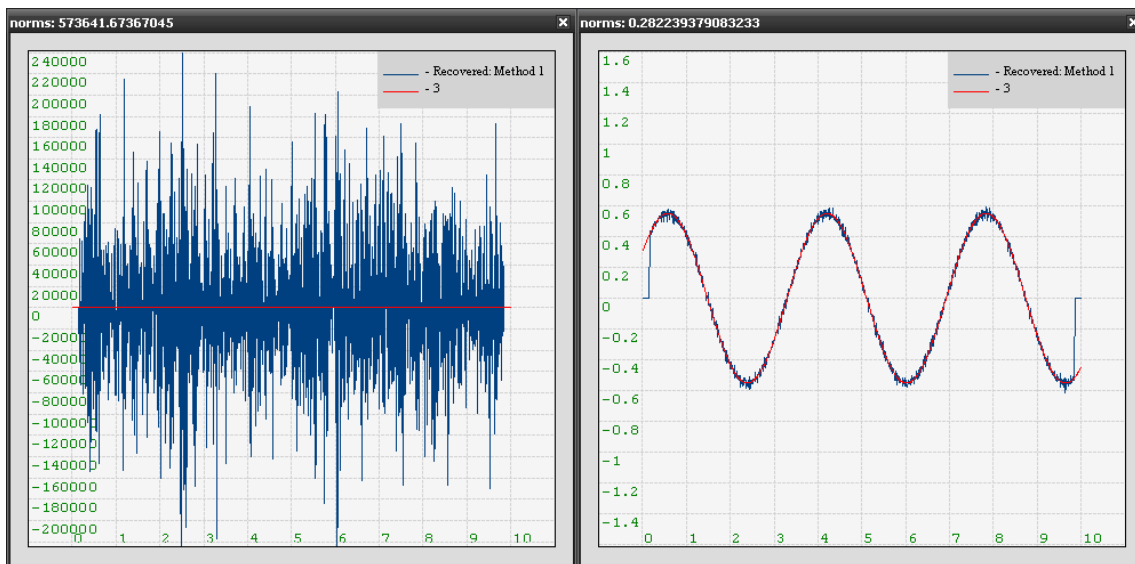


Figure 3: Non-smoothed estimation result of 2 methods.

Although after few steps of smoothing classical method gives acceptable results, it cannot be compared with results of the developed method, which perfectly fits the curve:



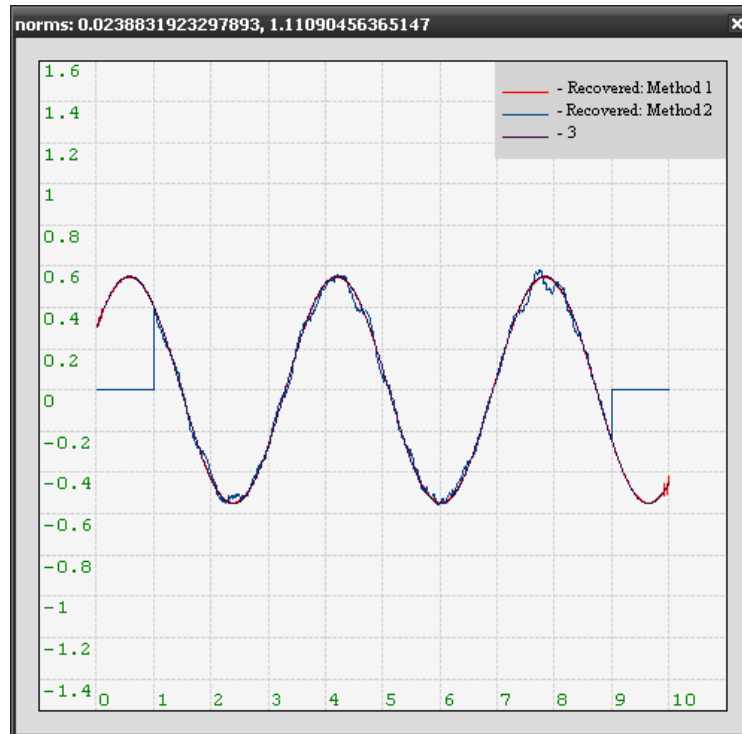


Figure 4: Smoothed estimation result of 2 methods.

In cases of larger measurement errors (variance=0.1) both methods give unacceptable results without smoothing, but this is what we have after smoothing (the best smoothing parameters for each method were selected):

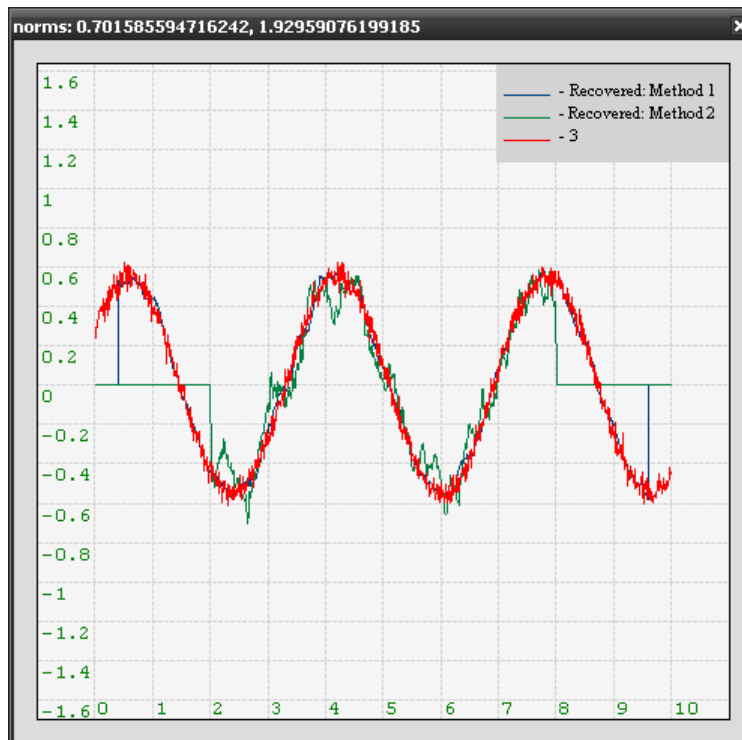


Figure 5: Smoothed estimation result of 2 methods

The order of derivative approximation method in (4) can also vary, influencing estimation result.

---

**Conclusion**

---

We have developed observability problem solving method, which is based on Pareto optimizations in the case of linear model in aerodynamics.

The method has been tested on generated data and showed the best results in comparison with classical methods.

As the problem has been investigated only in case of linear control, we shall continue work on considering wider range of models. Also we plan to run the method on real flight data and integrate received results with identification problem in case when some elements of matrix A are unknown.

---

**Bibliography**

---

[Zavorotnyy, 2004] Zavorotnyy, A. Decision of super-resolution measuring-calculating system modelling problem based on multicriterion optimization. – Bulletin of the University of Kiev, Series: Physics & Mathematics, 2004, #3, pp198-205.

[Zavorotnyy, 2006] Zavorotnyy, A. Development of mathematical methods and algorithms for decision of super-resolution measuring-calculating systems modelling problems. Manuscript, Kyiv, P.108., 2006

[Voronin, 1980] Voronin, A., Multicriterion optimization of dynamic control system. – Kibernetika, 1980, No. 4, pp. 56–68.

---

**Authors' Information**

---



**Andriy Zavorotnyy** – PhD, Researcher; TC science sector, theoretical cybernetics department, faculty of cybernetics, Kyiv National Taras Shevchenko University, Glushkov av. 2, build. 6, Kyiv-03127, Ukraine; e-mail: [zalbxod@mail.ru](mailto:zalbxod@mail.ru)

*Major Fields of Scientific Research: pareto-optimization, operator model of measuring-calculating system, fuzzy values*



**Veda Kasyanyuk** – PhD, Head of Science Sector, TC science sector, theoretical cybernetics department, faculty of cybernetics, Kyiv National Taras Shevchenko University, Glushkov av. 2, build. 6, Kyiv-03127, Ukraine; e-mail: [zalbxod@mail.ru](mailto:zalbxod@mail.ru)

*Major Fields of Scientific Research: measurements' reduction to calculations, calibration of unknown measurements' model, spline approximation*

## CYCLE ROUTES OPTIMIZATION FOR NOT FULL GRAPH

Anatoly Panishev, Anton Levchenko

**Abstract:** Searching for cycling routes of minimum cost is an urgent task of transport nets designing, traffic flows routing, etc. Each locality, transport node can be the vertex of the graph, and each cut line, linking localities, – the edge of the graph. The edge's cost represents distance, repair cost, channel bandwidth, circuit resistance, etc.

If closed route reaches all vertexes of the graph exactly once, it is called Hamiltonian cycle (HC), and the task of searching for minimum cost is the Hamiltonian Traveling Salesman's Problem (HTSP). HTSP is NP-hard and not always has a solution [Майнука, 1981]. Algorithms delivering optimum to HTSP are described as an exhaustive search reduction.

If restriction of visiting only once every vertex is taken off, then such task can be called Common Traveling Salesman Problem (CTSP). Besides completeness of the route is an additional restriction for this task. CTSP seems the generalization of HTSP [Бондаренко, 2004]. For every linked graph always the CTSP's set of solutions is not empty, and it allows using approximate algorithms and heuristics. The authors have no data about exact methods, bringing optimum of closed CTSP.

This article offers exact algorithm of CTSP solution, bringing the cycle of minimum cost, which includes all vertexes of the graph at least once and contains the less number of edges.

**Keywords:** TSP, Traveling Salesman Problem, Hamiltonian cycle, closed route, optimal route, exact algorithm.

**ACM Classification Keywords:** Algorithms, Theory.

---

### Introduction

Transport nets, electric chains are often presented as graph. Most methods come from the idea that graph is full. The ideas, offered in the article, allow using standard methods for arbitrary connected graphs.

One of the most famous graph problems is Traveling Salesman's Problem (TSP). We call Hamiltonian TSP (HTSP) task of finding minimal cost route visiting all graph's vertexes only once [Бондаренко, 2004]. If route can visit any vertex at least once we call such task Common TSP (CTSP).

---

### Algorithm's foundation

The graph  $H = (U, V)$  is given, where  $V$  – is the set of vertexes,  $U$  – is the set of edges, connecting these vertexes. Every edge  $\{i, j\} \in U$  has cost  $d_{ij} \in R_0^+$ , where  $R_0^+$  – is the set of real non-negative numbers. The task is to find in graph  $H$  minimal cost cycle, which reaches every vertex of the graph.

If  $H$  is linked, then every random pair of vertexes  $i, j \in V$  is connected by the set of simple chains  $A_{ij}$ , this set contains the chain  $\alpha_{ij}$  with the minimal total weight of included edges  $D(\alpha_{ij})$ . Matrix  $[D(\alpha_{ij})]_n$  specifies the full graph  $H_\alpha = (V, E)$ , with every edge  $\{i, j\} \in E$  responses to the chain  $\alpha_{ij}$  with weight  $D(\alpha_{ij})$  in the graph  $H$ .  $H_\alpha$  is full and meets triangle's condition

$$d_{ij} \leq d_{ik} + d_{kj}, \quad i \neq k \neq j \quad (1)$$

By analogy with (1) let's write following inequality, characterizing weight relationships between edges of the graph  $H$ , even if it is not full:

$$d_{ij} \leq D(\alpha_{ij}) \quad (2)$$

There are two possibilities: a) all edges of  $H$  meet inequality (2), b) at least one of the edges doesn't meet (2).

The graph  $H = (V, U)$  is assumed to be Hamiltonian;  $\tau$ ,  $T$  – are optimal solutions of HTSP and CTSP

accordingly, which are built in the  $H$ ,  $D(\tau)$  и  $C(T)$  – costs of these built solutions. Let's find the minimal cost cycle  $\sigma$  in the full graph  $H_\alpha$ .

*Predicate 1.* If inequality (2) holds for all edges  $\{i, j\} \in U$  of the Hamiltonian graph,  $H = (V, U)$  then  $T = \tau$ ,  $C(T) = D(\tau)$ .

*Demonstration.* HTSP solution  $\tau$  contains  $n$  edges of the set  $U$ . In case a) the edge  $\{i, j\}$  of the full graph  $H_\alpha$  has cost  $D(\alpha_{ij}) = d_{ij}$ , if  $\{i, j\} \in U$ , and  $D(\alpha_{ij}) \geq d_{ij}$  else. Implies, that in the  $H_\alpha$  exact solution of TSP  $\sigma$  matches  $\tau$ , and it's cost doesn't exceed the cost of random route which contains  $n$  and more edges, including  $T$ .  $\square$

*Predicate 2.* If at least for one edge  $\{i, j\}$  of the Hamiltonian graph,  $H = (V, U)$  inequality (2) doesn't hold, then  $C(T) \leq D(\sigma)$ .

*Demonstration.* In case b) in the full graph  $H_\alpha$  there is at least one edge  $\{i, j\}$ , which in the  $H$  has cost  $d_{ij} > D(\alpha_{ij})$ . If Hamiltonian cycle  $\sigma$  of the graph  $H_\alpha$  contains the edge  $\{i, j\}$ , then it's cost is greater than the cost of the corresponding route in the  $H$ , which contains chain  $\alpha_{ij}$  instead of the edge  $\{i, j\} \in U$ . If the cycle  $\sigma$  doesn't include edges, that break inequality (1), then it matches HTSP.  $\square$

The graph  $H = (V, U)$  is assumed not to be Hamiltonian. Then in any case a) or b) cost of the optimal TSP solution  $\sigma$  for the full graph  $H_\alpha = (V, E)$  is equal to cost of the CTSP's solution  $T$  for the graph  $H$ . So closed route  $T$  could be found by building Hamiltonian cycle  $\sigma$  in the graph  $H_\alpha$ , and replacing every edge  $\{i, j\} \in E$  by the chain  $\alpha_{ij}$  obtained from edges of the set  $U$ .

---

### Exact solution algorithm for CTSP

---

S0.  $H = (V, U)$  – linked weighted graph with the set of vertexes  $V$ ,  $|V| = n$ , and the set of edges  $U$ ,  $[d_{ij}]_n$  – the matrix of weights of the graph  $H$ , where if  $\{i, j\} \in U$  then  $d_{ij} \in R_0^+$ , else  $d_{ij} = \infty$ ;  $i, j = \overline{1, n}$ ,  $R_0^+$  – the set of real non-negative numbers.

S1. Build by Floyd's algorithm the matrix  $[\alpha_{ij}]_n$  of the shortest chains between all pairs of vertexes of the graph  $H$  and the matrix  $[D(\alpha_{ij})]_n$ , where every element  $(i, j)$  is equal to the cost  $D(\alpha_{ij})$  of the chain  $\alpha_{ij}$ ; matrices  $[\alpha_{ij}]_n$  and  $[D(\alpha_{ij})]_n$  sets the full weighted graph  $H_\alpha = (V, E)$ , where every edge  $\{i, j\}$  replaces the chain  $\alpha_{ij}$  in the graph  $H$ .

S2. Find minimal cost circuit  $\sigma$  in the graph  $H_\alpha$  by any known algorithm of metric TSP solution.

S3. Build the optimal solution of CTSP  $T$  replacing every edge  $\{i, j\}$  of circuit  $\sigma$  by the graph's  $H$  chain  $\alpha_{ij}$ .

---

### Modified Little's method

---

Graph  $H$  can contain several tours with the cost equal to optimal, but with different edges number, with the above described algorithm can select any of them. Proposed use at the step S2 modification of the classic Little's algorithm, which reaches optimum with less number of edges.

According to Little's method, building of the optimal solution occurs during branching with the purpose of partition of the set of feasible solutions to disjoint subsets. These subsets are represented as vertexes of the solutions tree

[Харани, 1973]. The root of the tree is the vertex  $\emptyset$ , denoting the set of all feasible TSP solutions. It is the initial vertex of branching.

Let  $D(\alpha_{ij})$  is value of the element  $(i, j)$  of the reduced matrix of costs of the initial graph,  $i, j = \overline{1, n}$ . Every matrix's element is represented as an arc in the oriented multigraph  $G_\alpha$ , where every two vertexes  $i$  and  $j$  are connected by the pairs of arcs  $(i, j)$  and  $(j, i)$ .

Branching starts with selection of the most appropriate for tour arc  $(k, l)$  of the multigraph  $G_\alpha$ . The set of feasible solutions is divided to two subsets: including the arc  $(k, l)$ , and not including that arc. These subsets vertexes can be marked as  $((k, l) \circ)$  and  $((\overline{k, l}) \circ)$  accordingly. For them lower bounds of TSP's costs  $\varphi((k, l) \circ)$  and  $\varphi((\overline{k, l}) \circ)$  are calculated. Reduced matrices are formed too. Among the terminal vertexes of the search tree the vertex with the lowest bound is determined – a branching vertex. According to the branching vertex the subset is divided to two subsets by arc, based on reduced matrix. For the corresponding vertexes reduced matrixes are formed and lower bounds calculated.

The way from the search tree's root to terminal vertex contains a part of a feasible TSP's solution consisting of arcs included in the process of branching. The method is considered completed when the next branching vertex includes all tour's arcs in multigraph  $G_\alpha$ .

The number of edges in a tour can be called length of tour. If at any step of the branching several terminal vertexes have the same minimum bound, then Little's method offers to choose any of them that do not guarantee a CTSP's solution with the lowest length. We introduce an additional parameter which is a lower bound for the length of the resulting tour for each vertex.

In the building of solutions tree the path from the root to terminal vertex contains the arcs included and not included in the partial solution. Let  $P_j = \{(v_{j_p}, v_{j_q})\}$  be the set of arcs included into the partial solution as a result of building of the path from root to terminal vertex  $j$ . Every arc  $(v_{j_p}, v_{j_q})$  of the set  $P_j$  in multigraph  $G_\alpha$  is an edge  $(v_p, v_q)$ , in the graph  $H$  with corresponding shortest chain  $\alpha_{pq}$ , which connects vertexes  $p$  and  $q$  and contains  $l(\alpha_{pq})$  edges. Then the partial solution for vertex  $j$  includes  $I_j = \sum_{(v_{j_p}, v_{j_q}) \in P_j} l(\alpha_{pq})$  edges of the graph  $H$ .

Reduced matrix  $[d']_j$  of the vertex  $j$  contains at least one zero element in every column and row.  $d'_{ij} = 0$  means the arc  $(i, j)$  of multigraph  $G_\alpha$  is the most likely candidate for inclusion in the desired tour. From  $[d']_j$  columns  $\{j_p\}$  and rows  $\{j_q\}$  are deleted corresponding to arcs  $\{(v_{j_p}, v_{j_q})\}$ , included to partial solution  $P_j$ . Zeros in the same row (column) of the matrix  $[d']_j$  correspond to arcs of  $G_\alpha$ , starting or ending in the same vertex.

Let's build matrix  $[d'']_j$  in such a way that if  $d'_{ij} \neq 0$ , then  $d''_{ij} = \infty$ , else  $d''_{ij} = l(\alpha_{ij})$ . Obviously, adding  $G_\alpha$  to tour arc  $\{i, j\}$  with  $d'_{ij} = 0$ , results to the fact that tour in the graph  $H$  increases by  $d''_{ij}$  edges. Let  $\tau''_j$  is a permutation of columns of matrix  $[d'']_j$ , which is solution of Assignment Problem (AP). The cost of this solution  $C(\tau''_j)$  is the sum of diagonal elements of permuted matrix and this sum is minimal. Corresponding to diagonal

matrix's elements edges don't start or finish at the same vertex. Consequently  $C(\tau_j^n)$  is the lower bound for the number of edges of graph  $H$ , which can be included to the tour in the further branching of the search tree.

Let every terminal vertex  $j$ , is characterized by value:

$$L_j = I_j - |P_j| + C(\tau_j^n) \quad (3)$$

in addition to lower bound  $\varphi(j)$ . If lower bound of several terminal vertexes is equal then we choose the vertex with minimal value of (3) that facilitates the choice of the optimal solution with less length.

---

### Graph's topology and algorithm's productivity

---

Graph of the actual transport network carries information to assess whether it is connected. The three components of the graph describing its connectivity are the articulation point, bridge, and a subset of the pendant or terminal nodes.

If deletion of vertex transforms a connected graph into a disconnected, then it is called the point of articulation (cutpoint), and an edge with the same property - a bridge. A connected, non-empty, having no cutpoints subgraph of graph  $H$  is called a block. Cutpoint is a common vertex of several blocks [Харари, 1973]. Non-empty subset of terminal nodes of graph  $H$  creates subgraph  $H'$  in the form of forest [Гаращенко, 2007]. Every tree of the forest corresponds to root node, connecting it with the part of the graph  $H$ , which doesn't include edges and other vertexes of the tree. Obviously, the each tree's root vertex is a point of articulation, and therefore, every tree of the forest is a block where all vertexes are the points of articulation, except for the hanging ones.

The above mentioned algorithm of CTSP's exact solution permits improvement, reducing the work time of the branch and bound procedure for a graph containing blocks of trees and bridges. The improvement mechanism is based on demonstrable fact.

*Predicate 3.* Any CSTP's solution for a tree has a value equal to the double sum of weights of tree's edges.

Let graph  $H$  contain blocks built from bridges and trees. First we select all blocks-trees in  $H$  and build for each of them closed route starting in route vertex and visiting every vertex, and passing every edge twice.

Each route planning, we note the root vertex. Then we select in graph  $H$  all bridges and mark the articulation point of each bridge. Obviously, the cost of a closed path on the bridge is equal to the double weight of the edges, representing the bridge. The sum of the values of the routes built for the selected objects is a constant component of the cost of any CTSP's solution in graph  $H$ .

Let's study the subgraph obtained by removing from the graph  $H$  all bridges and all trees except root vertexes. The subgraph is not linked if the graph  $H$  contains bridges. The connected components of subgraphs together contain all the marked vertexes. The component could be an articulation point, a block or a decomposable subgraph, i.e. subgraph containing the articulation points, other than marked. Then we construct the salesman's route for each component, which begins and ends at the marked nodes incidental to the bridge. CTSP's solution for graph  $H$  is the result of union all successive routes in the marked vertexes.

To select the trees of the forest  $H'$  apply version of the algorithm proposed in [Гаращенко, 2007]. Mark as  $V'$  and  $U'$  sets of vertexes and matrices of the forest  $H'(V', U')$ ,  $K$  - set of root vertexes,  $K \subset V'$ . The following algorithm constructs all the trees of the forest  $H'$  in graph  $H$  and defines set of root vertexes  $K$ .

S0.  $H = (V, U)$  - linked graph where  $V$  - set of vertexes,  $U$  - set of edges  $u = \{i, j\}$ ,  $|V| = n$ ; vertexes of graph  $H$  placed in non-decreasing order of their degrees:  $\deg 1 \leq \deg 2 \leq \dots \leq \deg n$ ;  $K = \emptyset$ ,  $V' = \emptyset$ ,  $U' = \emptyset$ .

S1. If  $\deg 1 > 1$ , then end: graph  $H$  doesn't contents graph  $H'$ , else  $i = 1$ .

S2. For edge  $u = \{i, j\}$  let  $\deg i = 1$ ,  $\deg j = \deg j - 1$ ,  $V = V - \{j\}$ ,  $V' = V' \cup \{i, j\}$ ,  $U' = U' \cup \{u\}$ ;  $i = i + 1$ .

S3. If  $i = n - 1$ , then  $V' = V$ ,  $U' = U$ , end: graph  $H$  is a tree.

S4. If  $\text{deg } i = 1$ , then go to S2.

S5.  $K = V' \cap V$ , to build for every root vertex from  $K$  tree of forest  $H'$ .

Time of the algorithm is obviously commensurable with the time of ordering vertexes' degrees of the graph  $H$ , evaluated, as it is well known by value  $O(n \log n)$ .

Every edge  $\{v, w\}$  of the tree  $H'_k$ ,  $k = \overline{1, |K|}$ , generally speaking, is a bridge, the removal of which leads to a non-connected spanning subgraph of graph  $H$ , not containing  $\{v, w\}$ .

Let's consider the subgraph  $\langle S \rangle$  of the subgraph  $H$ , generated by a subset of vertexes  $S = (V - V') \cup K$ . To find and select all the bridges in the subgraph  $\langle S \rangle$  depth-first search algorithm is used that determines the set of vertexes  $S$  all articulation points during  $O(|S| + |V| - |V'|)$  [Рейнгольд, 1980]. If a pair of cutpoints is connected by an edge deletion, which increases the number of connected components of the subgraph  $S$ , then it creates bridge  $M_m$ . Set of  $s$  bridges generates  $s + 1$  connected components  $H''_l$  of subgraph  $\langle S \rangle$ . Closed route visiting all vertexes of the component  $H''_l$ ,  $l = \overline{1, s + 1}$ , and delivering the minimum sum of weights of edges, is a part of traveling salesman's tour  $\tau$  in graph  $H$ .

We show how to unite to CSTP's solution  $\tau$  routes  $\tau'_k$ ,  $e_m$ ,  $\tau''_l$ , which are built for trees  $H'_k$  of the forest  $H'$ ,  $k = \overline{1, |K|}$ , bridges  $M_m$ ,  $m = \overline{1, s}$ , and connected components  $H''_l$  of subgraph  $\langle S \rangle$ ,  $l = \overline{1, s + 1}$ .

Among all the components  $H''_l$  always exists such that contains exactly one vertex  $V_u$  of the bridge. It is defined as the first in the set of components  $H''_l$ ,  $l = \overline{1, s + 1}$ . The vertex  $V_u$  is marked as the starting and ending point of a salesman's route  $\tau$ . If  $s > 2$ , then the part of components  $H''_l$ ,  $l = \overline{2, s + 1}$ , is interconnected by several bridges. In such components there is one vertex of each bridge.

Let  $\tau''_l = (w, \dots, a, \dots, b, \dots, w)$  be the salesman's tour built for  $H''_l$ ,  $l = \overline{1, s + 1}$ , where  $w$  – bridge's vertex;  $w = v$  if  $l = 1$ . Suppose it contains vertexes from the set of root vertexes  $K$ . Then we execute in  $\tau''_l$  replacement of every vertex  $a \in K$  by the route  $(a, \dots, a)$  for the tree  $H'_k$  with ending vertex  $a$ . We mark all vertexes in every route  $\tau''_l$ , incidental to bridges.

Let's build graph  $(L, M)$ , in which to every vertex  $l \in L$ ,  $l = \overline{1, s + 1}$ , is assigned to the route  $\tau''_l$ , and to every edge  $\{i, j\} \in M$  – a bridge  $M_m$ ,  $m = \overline{1, s}$ . In graph  $(L, M)$  a pair of vertexes  $i$  and  $j$  creates an edge  $\{i, j\}$ , if vertexes  $p \in \tau''_i$  and  $q \in \tau''_j$  are connected by bridge  $\{p, q\}$ ,  $i \neq j$ . Graph  $(L, M)$  is a tree, since by construction it is linked, and  $|L| = |M| - 1$ . We assume vertex 1 of the tree  $(L, M)$  as a root vertex. To salesman's tour  $\tau$  every tour is assigned including all tree's vertexes, which starts and ends in the vertex 1 and runs twice along every edge from  $M$ .

All details of uniting to tour  $\tau$  of sets of routes  $\{\tau''_l | l = \overline{1, s + 1}\}$ , linked by bridges  $M_m$ ,  $m = \overline{1, s}$ , can be presented on the basis of Fig. 1. First  $\tau$  includes tour  $\tau''_1 = (v, \dots, v)$  and the bridge  $\{v, w\}$ , linking  $\tau''_1$  with route  $\tau''_2 = (w, \dots, x, \dots, y, \dots, w)$ . As  $x$  and  $y$  – marked vertexes, i.e. vertexes of bridges  $\{x, a\}$  и  $\{y, b\}$ , then  $\tau$  runs along the part  $(w, \dots, x)$  of tour  $\tau''_2$ , edge  $\{x, a\}$ , route  $\tau''_3 = \{a, \dots, a\}$  and edge  $\{a, x\}$ . Then it includes the part  $(x, \dots, y)$  of the route  $\tau''_2$ , edge  $\{y, b\}$ , route  $\tau''_4 = (b, \dots, b)$ , edge  $\{b, y\}$ , one more part  $(y, \dots, w)$  of the route  $\tau''_2$  and the edge  $\{w, v\}$ . Thus the tour  $\tau = (v, \dots, v, w, \dots, x, a, \dots, a, x, \dots, y, b, \dots, b, y, \dots, w, v)$  is built. Obviously, the time to unite all the selected subgraphs of the graph  $H$  is limited by value  $O(|V|)$ .

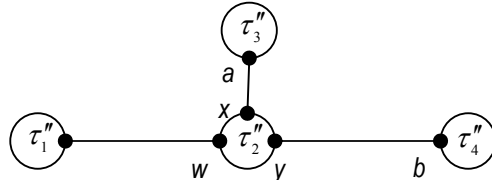


Fig. 1. The tree  $(L, M)$  with distinguished vertices

Salesman's tour's  $\tau$  cost equals to  $C(\tau) = \sum_{k=1}^{|K|} C(\tau'_k) + 2 \sum_{m=1}^s C(M_m) + \sum_{l=1}^{s+1} C(\tau''_l)$ , where the first two terms are constants for a given graph  $H$ . The cost of subgraph's  $\langle S \rangle$  components, which is represented by one vertex, is equal to 0.

Example 1. At the Fig 2. a connected weighted graph  $H$  is shown. Closed salesman's tour  $\tau$  is required to be built for it.

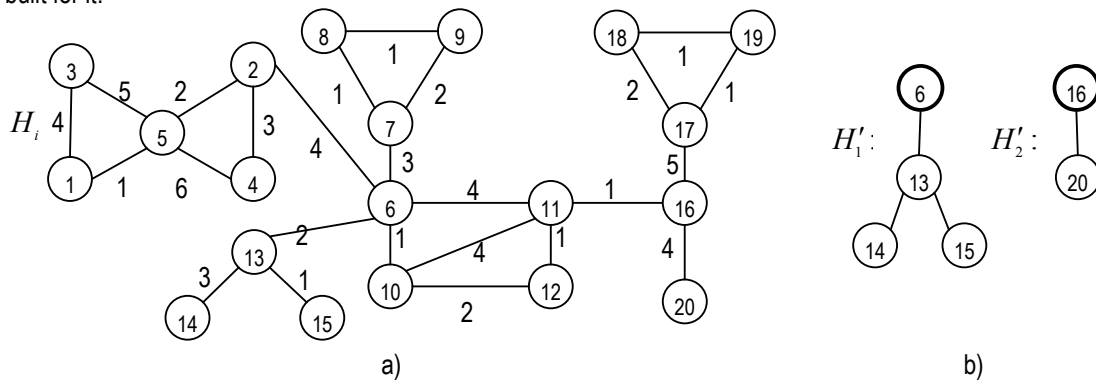


Fig. 2 a) original graph; b) trees of the forest  $H'$ .

First the forest's  $H'$  building algorithm is performed, which consists of two stages. In the first stage it sorts all graph's  $H$  vertices by their degrees in non-decreasing: deg 14=1, deg 15=1, deg 20=1, deg 1=2, deg 3=2, deg 4=2, deg 8=2, deg 9=2, deg 12=2, deg 18=2, deg 19=2, deg 2=3, deg 13=3, deg 7=3, deg 10=3, deg 16=3, deg 17=3, deg 5=4, deg 11=4, deg 6=5. In the second stage algorithm finds the set of forest's vertices  $V = \{14, 15, 13, 6, 20, 16\}$ , subset of root vertices  $K = \{6, 16\}$  and builds  $n$  trees  $H'_1, H'_2$  (Fig. 2, b). Trees' root vertices at Fig. 2 b) are marked by thicker lines of circles.

Generated by subset  $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 19\}$  subgraph  $\langle S \rangle$  is shown at fig 3 a).

Finding cutpoints algorithm finds there are 7 such cutpoints in  $\langle S \rangle$ : 5, 2, 6, 7, 11, 16, 17. Vertices 2, 6, 7, 11, 16, 17 from them create bridges  $M_1 = \{2, 6\}$ ,  $M_2 = \{6, 7\}$ ,  $M_3 = \{11, 16\}$ . Deletion of bridges makes 5 connected components of subgraph  $\langle S \rangle$   $H''_1, H''_2, H''_3, H''_4, H''_5$  (Fig. 3, b).

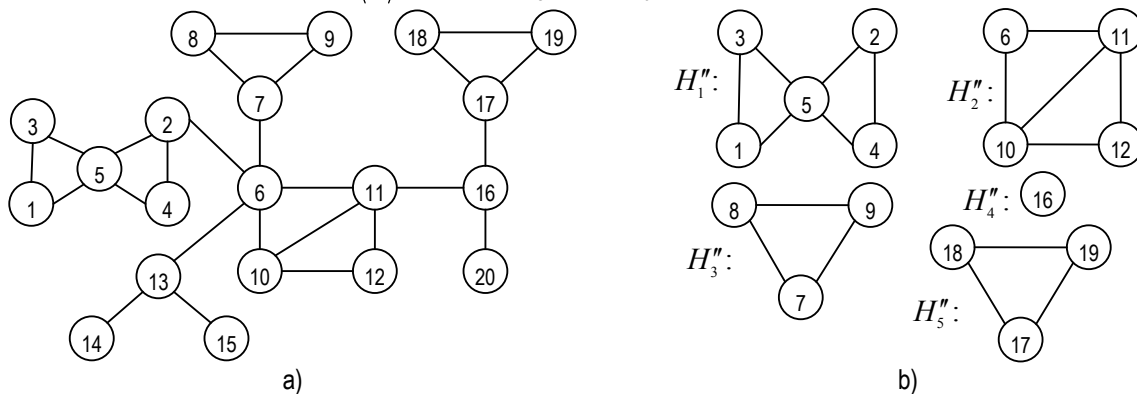


Fig. 3 a) the induced subgraph  $\langle S \rangle$ ; b) subgraph's  $\langle S \rangle$  connected components.



We build closed salesman's routes, starting and ending at the marked vertexes, for each tree and each connected component of subgraph  $\langle S \rangle$ :  $\tau'_1 = (6, 13, 14, 13, 15, 13, 6)$ ,  $\tau'_2 = (16, 20, 16)$ ,  $\tau''_1 = (2, 4, 2, 5, 3, 1, 5, 2)$ ,  $\tau''_2 = (6, 10, 12, 11, 6)$ ,  $\tau''_3 = (7, 8, 9, 7)$ ,  $\tau''_4 = (16)$ ,  $\tau''_5 = (17, 18, 19, 17)$ . As a result of route  $\tau'_1$  unites with route  $\tau'_2$  we get the tour  $\tau''_2 = (6, 13, 14, 13, 15, 13, 6, 10, 12, 11, 6)$ . Union  $\tau''_2$  with  $\tau''_4$  gives the rout  $\tau''_4 = (16, 20, 16)$ .

The tree  $(L, M)$ , built for subgraph  $\langle S \rangle$  of graph  $H$ , is presented at Fig 4.

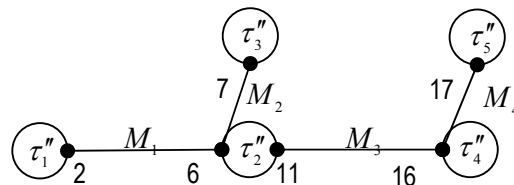


Fig. 4. The corresponding to subgraph  $\langle S \rangle$  of graph  $H$  tree  $(L, M)$ .

Select any tour of the tree's  $(L, M)$  vertexes, which starts and ends in vertex 1. Assume that it is  $(1, 2, 3, 2, 4, 5, 4, 2, 1)$ . Build salesman's tour  $\tau$  for graph  $H$   $\tau = (2, 4, 2, 5, 3, 1, 5, 2, 6, 13, 14, 13, 15, 13, 6, 7, 8, 9, 7, 6, 11, 16, 20, 16, 17, 18, 19, 17, 16, 11, 12, 10, 6, 2)$ . In  $\tau$  bridge's vertexes are marked bold, and rout  $\tau''_2$  is presented by two parts  $(6, 11)$  and  $(11, 12, 10, 6)$ , which are doesn't following one after the other. The cost of such CTSP's solution for the graph  $H$  is  $C(\tau) = \sum_{k=1}^2 C(\tau'_k) + 2 \sum_{m=1}^4 C(M_m) + \sum_{l=1}^5 C(\tau_l) = (2+3+3+1+1+2) + (4+4) + 2(4+3+1+5) + (3+3+2+1+4+5+2) + (1+2+ +1) + (2+2+1) = 75$ .  $\square$

## Conclusion

The exact algorithm of CTSP solution is offered. The modification of Little's method is also offered which chooses among several optimal solutions the one with the least number of graphs. The procedure of branching for graphs with the determined topology can be more effective. Some features of his topology are described in the paper.

## Bibliography

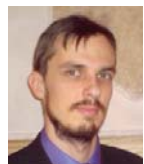
- [Майника, 1981] Майника Э. Алгоритмы оптимизации на сетях и графах. – М.: Мир, 1981. – 323 с.
- [Бондаренко, 2004] Бондаренко М.Ф., Белоус Н.В., Руткас А.Г. Компьютерная дискретная математика. – Харьков: «Компания СМИТ», 2004. – 476 с.
- [Харари, 1973] Харари Ф. Теория графов. – М.: Мир, 1973. – 300 с.
- [Гаращенко, 2007] Гаращенко И.В., Панишев А.В. Об одном классе задач построения остовного дерева в неориентированном взвешенном графе // Искусственный интеллект. – Донецк, 2007. – Вып. 3. – С. 486-493
- [Рейнгольд, 1980] Рейнгольд Э., Нивергельт Ю, Део Н. Комбинаторные алгоритмы. Теория и практика. – М.: Мир, 1980. – 476 с.

## Authors' Information



**Anatoliy Panishev** – Professor, Doctor of Technology, Chief of the Chair of Informatics and Computer Modeling ZSTU; Chernyakhovsky str. 103, 10005 Zhitomir, Ukraine; e-mail: [irinagar@rambler.ru](mailto:irinagar@rambler.ru)

Major Fields of Scientific Research: Combinatorial optimization, Graph theory



**Anton Levchenko** – Assistant, Chair of Informatics and Computer Modeling ZSTU; Chernyakhovsky str. 103, 10005 Zhitomir, Ukraine ; e-mail: [dear\\_anton@yahoo.com](mailto:dear_anton@yahoo.com)

Major Fields of Scientific Research: Combinatorial optimization, Software technologies

## MODIFIED BRANCH AND BOUND ALGORITHM FOR SOLVING THE HAMILTONIAN RURAL POSTMAN PROBLEM

Andriy Morozov, Anatoliy Panishev

**Abstract:** In this paper the Hamiltonian Rural Postman Problem is generalization of the Hamiltonian Travelling Salesman Problem is described. The offered modification of a classical method (Little's method) allows to find exact solution of the Hamilton Rural Postman Problem or to detect that the task is unsolvable.

**Keywords:** Hamiltonian Rural Postman Problem, Hamiltonian cycle, branch and bound algorithm, Hamiltonian Traveling Salesman Problem

**ACM Classification Keywords:** G.2.2 - Mathematics of Computing - Discrete Mathematics - Graph Theory - Path and circuit problems

### Introduction

Many results are saved up in research of the Travelling Salesman Problem. They cover the important problems of algorithm design and improvement of methods of combinatorial optimization and their application.

Algorithms with indicators of efficiency which are applicable in real situations aren't known for each applied problem of type of the Travelling Salesman Problem. One of such problems is Rural Postman Problem, formulated as follows [1].

Let  $H=(V, U)$  is connected weighed graph,  $V$  – set of vertices,  $|V| = n$ ,  $U$  - set of edges. Each edge  $\{i, j\} \in U$  has the weight  $d_{ij} \in Z_0^+$ ,  $i \neq j$ ,  $i, j = \overline{1, n}$ ,  $Z_0^+$  - set of non-negative numbers. The symmetric matrix of weights  $[d_{ij}]_n$  completely defines weighted graph  $H$ . At this matrix: if  $\{i, j\} \in U$  then  $d_{ij} \in Z_0^+$ , else  $d_{ij} = \infty$ ,  $i \neq j$ ,  $i, j = \overline{1, n}$ ,  $d_{ii} = \infty$ ,  $i, j = \overline{1, n}$ . The nonempty subset of edges  $R \subseteq U$  is given. It is required to find a cycle which includes each edge from  $R$  and has minimum sum of weights of edges.

Let's designate the Hamiltonian cycle of graph  $H$  which is passing on all edges of set  $R$  as  $z(R)$ . Let's name as the Hamilton Rural Postman Problem (HRPP) the problem which consists in determination of the Hamiltonian cycle  $z^*(R)$  minimizing a functional

$$C(z(R)) = \sum_{\{k,l\} \in z(R)} d_{kl} . \quad (1)$$

Interest in solving the HRPP arises when it is required to find an annular route on a transport network of a city or the district, modeled by graph  $H = (V, U)$ . To each point of departure (arrival) of a network corresponds vertex  $i \in V, |V| = n$ , and to each edge  $\{i, j\} \in U$  corresponds a segment of a road between pair of adjacent points  $i$  and  $j$ . Edge  $\{i, j\}$  is characterized by weight (cost)  $d_{ij}$ . It is equal to expenses for vehicle movement from  $i$  to  $j$  or from  $j$  to  $i$ .

### Algorithm

HRPP is NP-complete, because it is a NP-complete Hamiltonian Travelling Salesman Problem (HTSP) in case when  $R = \emptyset$ .

In [2] there is the algorithm which correctly finds solution of HTSP if graph  $H$  is Hamiltonian, or detects its unsolvability. Basis of the offered algorithm is a method of branch and bound scheme. It is fulfilled after check of sufficient conditions of unsolvability HRPP. Clearly, that the complexity of such check should be limited by a polynomial from a problem size.

Direct application of branch and bound algorithm from [2] does not allow to solve HRPP. Inclusion of a subset of edges  $R \neq \emptyset$  in a required Hamilton cycle turned out so strong restriction that demands other approach to the organization of branching and an evaluation of the lower bound for  $C(z^*(R))$ .

Obviously, if graph  $H$  contains suspended vertices HTSP and HRPP has no solutions. Suspended vertices in graph  $H$  are finding with complexity  $O(|V|)$ . The problems are unsolvable, when graph  $H$  has a concatenation point [2, 3]. It is required  $O(|V|+|U|)$  elementary operations to define whether graph  $H$  contains a concatenation point [3].

It is easy to see, that HRPP is unsolvable, if in graph  $H$  a) the subset of edges of set  $R$  forms nonhamiltonian cycle; b) there is a vertex, which is incident three or more edges from  $R$ . Therefore, graph  $H$  in which the set of edges  $R$  does not form a collection of vertex-not crossed chains, does not contain cycle  $z(R)$ . Time limited in size  $O(|V|+|U|)$  suffice to check the conditions.

It is known from [4], that HRPP can be transformed to the same task on graph  $H = (V, U)$  in which a) degrees of all vertices are higher 2, b) the set of edges  $R$ , which is contain in a required Hamiltonian cycle, organizes matching  $R$ .

Let's notice, that  $|R|$  limits a solution space so, that it can be empty. Let's name admissible solution  $z(R)$  of HRPP by tour.

Search the solution of the task starts with conversion a matrix of costs of graph  $H$  to the reduced matrix [5]. The reason is that there is minimum element  $\alpha_i = \min_j d_{ij}$ ,  $i = \overline{1, n}$  in line  $i$  which is subtracted from each element of this line. Then in the column  $j$  minimum element  $\beta_j = \min_i d_{ij}$ ,  $j = \overline{1, n}$  which is subtracted from each element of this column is retrieved. Elements  $\alpha_i$  and  $\beta_j$  are named as reducing coefficients. From the reduced matrix  $[d_{ij}]_n$  the lower bound of cost of the required solution is searched:

$$\varphi(R) = \sum_{\{i,j\} \in R} \min\{d_{ij}, d_{ji}\} + \gamma, \quad (2)$$

$$\text{where } \gamma = \sum_{i=1}^n \alpha_i + \sum_{j=1}^n \beta_j.$$

Generally, the reduced matrix which defines set of all solutions HRPP is not symmetric. To this matrix the weighed multigraph  $H' = (V, U')$ , in which vertices  $i$  and  $j$  are linked by pair arcs  $(i, j)$  and  $(j, i)$  if in graph  $H = (V, U)$  they are linked by edge  $\{i, j\}$ , corresponds one-to-one. Thus, each edge  $\{i, j\} \in R$  of graph  $H$  is presented in multigraph  $H'$  by two arcs  $(i, j) \in R'$ ,  $(j, i) \in R'$ ,  $|U'| = 2|U|$ ,  $|R'| = 2|R|$ .

From the reduced matrix we select elements  $(i, j)$  and  $(j, i) \in R'$  for which  $\Delta_{ij} = \min\{d_{ij}, d_{ji}\} > 0$  and let

$$d_{ij} = d_{ij} - \min\{d_{ij}, d_{ji}\} = d_{ij} - \Delta_{ij}, \quad d_{ji} = d_{ji} - \min\{d_{ij}, d_{ji}\} = d_{ji} - \Delta_{ij}.$$

Let's name obtained matrix  $[d_{ij}]_n$  completely reduced. The further operations will be fulfilled by means of this matrix.

Conversion of the matrix of costs of graph  $H$  to completely reduced matrix has the argued substantiation. HRPP for completely reduced matrix consists in construction the Hamiltonian cycle or detour the minimum cost which includes exactly one arc from each pair  $(i, j)$ ,  $(j, i) \in R'$ , containing at least one arc with zero weight. Thus, the list of arcs of the zero weight which are candidates for inclusion in optimal solution is defined on matrix  $[d_{ij}]_n$ .

The explained reasons open possibility to adapt the classical algorithm of branch and bound described in [6], for

searching the solution of the HRPP. The mode of branching of admissible solutions  $z(R)$  keeps within the known scheme of construction of a binary search tree [6].

Let's put in correspondence to the root of a tree of search  $\{z(R)\}$  completely reduced matrix  $[d_{ij}]_n$  with bound  $\varphi(R)$ , define the arc of multigraph  $H'$  which initiates branching. To realize it, as well as at [6], each element  $(p, q)$  in  $[d_{ij}]_n$ , if  $d_{pq} = 0$ , let's estimate as value

$$\gamma(p, q) = \min_{i \neq q} d_{pi} + \min_{j \neq p} d_{jq}, \quad (3)$$

also find the element  $(p, q)$  which has the greatest value

$$\Gamma(k, l) = \max\{\gamma(p, q) \mid d_{pq} = 0\}, \quad (4)$$

In multigraph  $H'$ , the arc  $(p, q)$  corresponds to element  $(p, q)$ . This arc initiates a partition of set of all detours to two subsets and, in the conditions of HRPP, generates two cases:  $\{k, l\} \notin R$  and  $\{k, l\} \in R$ .

In case of  $\{k, l\} \notin R$  the set of all solutions of the problem is divided into subsets  $\{\{k, l\} \notin R\}$  and  $\{\{\overline{k, l}\}\}$ . The first subset includes all detours which contain arc  $(k, l)$ , the another one includes all detours which do not contain this arc.

The matrix by which the low bound  $\varphi(\{k, l\} \in R')$  of cost of all detours of set  $\{\{k, l\} \notin R'\}$  is calculated, is provided as in [6] according to the following rule.

If the set  $R$  contains the edge  $\{x, k\}$ , the required detour together with the arc  $(k, l)$  joins the arc  $(x, k)$ . Similarly, if the set  $R$  contains the edge  $\{y, l\}$ , the arc  $(k, l)$  joins the arc  $(l, y)$ . Inclusion of the arc  $(x, k)$  or the arc  $(l, y)$  in a subset of solutions  $\{\{k, l\} \notin R'\}$  means, that the matrix which defines it, does not contain the line  $k$  and the column  $l$ , as well as the line and the column, numbers of which are the beginning and the end of the joined arc. In case of  $\{x, k\}, \{y, l\} \in R$ , the arcs  $(x, k)$  and  $(l, y)$  join to arc  $(k, l)$ , and in the matrix, which defines the set  $\{\{k, l\} \notin R'\}$ , the lines  $x, k, l$  and the columns  $k, l, y$  exclude.

Let's designate the lower bound of branching vertex as  $\varphi$ . For an arc  $(k, l) \notin R'$  which initiates branching, let

$$\mu_k = \begin{cases} d_{xk}, & \text{if } \{x, k\} \in R, \\ 0, & \text{else.} \end{cases}, \quad \mu_l = \begin{cases} d_{ly}, & \text{if } \{l, y\} \in R, \\ 0, & \text{else.} \end{cases}$$

Here  $d_{xk}, d_{ly}$  - the elements of a matrix which corresponds to the bound  $\varphi$ . In case  $\varphi = \varphi(R)$  they are the elements of the completely reduced matrix  $[d_{ij}]_n$ . Then, the cost of all detours of the set  $\{\{k, l\} \notin R'\}$  is bounded below by the value

$$\varphi(\{k, l\} \notin R) = \varphi + \mu_k + \mu_l + \sum \alpha_i' + \sum \beta_j', \quad (5)$$

where  $\alpha_i'$  and  $\beta_j'$  - the reduction factors. Those coefficients are obtained as a result of transformation of a matrix, which to the bound  $\varphi$  corresponds, to a matrix, which defines the set  $\{\{k, l\} \notin R'\}$ .

The matrix which defines the set  $(\overline{k, l})$ , and the bound  $\varphi(\overline{k, l})$  are presented by the value  $\varphi(\overline{k, l}) = \varphi + S(k, l)$ , where  $S(k, l)$  is the sum of reduction factors which is obtained as a result of reduction of a matrix of costs.

Let's describe a case  $(k, l) \in R$ . Solution space HRPP  $\{z(R)\}$  is divided into two subsets  $\{(k, l)\}$  and  $\{(l, k)\}$ . The first subset contains all detours which include the arc  $(k, l)$ , the second one - all detours which include the arc  $(l, k)$ .

To the branching vertex boundary  $\varphi$  and matrix  $D$  which generates arc  $(k, l)$  or  $(l, k)$ ,  $\{k, l\} \in R$ , with a maximum estimation let correspond. The matrix which defines subset  $\{(k, l)\}$ , is a result of elimination from  $D$  a line  $k$  and a column  $l$ , appropriations the value  $\infty$  to element  $d_{lk}$  and reduction of the obtained shorted matrix. Similarly, for determination a matrix which defines subset  $\{(l, k)\}$ , it is necessary in  $D$  to suppose  $d_{kl} = \infty$ , to remove a line  $l$  and a column  $k$  and to fulfill the reduction of the obtained shorted matrix. The lower boundaries of cost of detours for subset  $\{(k, l)\}$  and  $\{(l, k)\}$  are calculated as follows:

$$\varphi(k, l) = \varphi + d_{kl} + \sum_{i \neq k} \alpha_i' + \sum_{j \neq l} \beta_j', \quad (6)$$

$$\varphi(l, k) = \varphi + d_{lk} + \sum_{i \neq k} \alpha_i' + \sum_{j \neq l} \beta_j', \quad (7)$$

where  $\alpha_i', \beta_j'$  - reduction factors which are obtained as a result of transformation of a matrix  $D$ .

The situation when either  $d_{kl} = \infty$ , or  $d_{lk} = \infty$  in  $D$  is possible. It follows from (1), that if  $d_{kl} = \infty$ ,  $\{(k, l)\} = \emptyset$ , if  $d_{lk} = \infty$ ,  $\{(l, k)\} = \emptyset$ .

Modified Little's method of solving the Hamiltonian Rural Postman Problem has the following appearance.

0.  $H = (V, U)$  - the nonoriented weighed graph with degrees of vertices higher than 2, presented by a matrix of costs with power  $n = |V|$ ;  $R$  - matching of graph  $H$  which contains in required Hamiltonian cycle  $z^*(R)$  of the minimum cost. Let  $C^* = \infty$ .

1. The matrix of costs of graph  $H$  is transformed to the reduced matrix. From it, the lower bound  $\varphi(R)$  of costs for all solutions of a problem is calculated. From the reduced matrix completely reduced matrix  $D = [d_{ij}]_n$  which corresponds to the oriented weighed multigraph  $H' = (V, U')$ ,  $|U'| = 2|U|$ ,  $|R'| = 2|R|$  is defined. In  $H'$  it is required to find the Hamiltonian cycle (detour) which contains exactly the one arc from each pair of arcs  $(i, j)$ ,  $(j, i) \in R'$  and has the minimum cost.

2. In matrix  $D$  under formulas (3) and (4) an element  $(k, l)$  is calculated. In a multigraph it is presented by an arc  $(k, l)$  which generates the branching vertex in a search tree.

3. If  $\{k, l\} \notin R$ , then go to step 7.

4. If  $d_{lk} = \infty$ , then a subset of detour  $\{(l, k)\}, \{k, l\} \in R$ , which contains an arc  $(l, k)$  is empty; go to step 6.

5. To define the subset of detours  $\{(l, k)\}, \{k, l\} \in R$  which contain an arc  $(l, k)$ , the reduced matrix in which  $d_{kl} = \infty$  and with excluded line  $l$  and the column  $k$  is considered. After matrix reduction under the formula (7), the lower boundary  $\varphi(l, k)$  of cost of arcs of all detours of subset  $\{(l, k)\}$  is calculated. Subset  $\{(l, k)\}$  generates in a search tree the current active vertex  $\{(l, k)\}$ . If  $C^* > \varphi(l, k)$  then vertex  $\{(l, k)\}$  with reduced matrix  $D$  and bound  $\varphi(l, k)$  joins branching vertex, else it is not considered further. Go to step 9.

6. To determinate the subset of detours  $\{(k, l)\}, \{k, l\} \in R$  which contain an arc  $(k, l)$ , the reduced matrix in which  $d_{kl} = \infty$  and both the line  $k$  and the column  $l$  are excluded is considered. Under the formula (6) the lower

bound  $\varphi(k, l)$  of cost of all detours of subset  $\{(k, l)\}$  which initiates in a searching tree the current active vertex  $\{(k, l)\}$  is calculated. If  $C^* > \varphi(k, l)$ , then a vertex  $\{(k, l)\}$  with a reduced matrix  $D$  and a bound  $\varphi(k, l)$  joins the branching vertex, otherwise the vertex is excluded from the further reviewing; go to step 9.

7. In determination of the subset of detours  $\{(\overline{k}, l)\}, \{(k, l) \notin R$  which is not containing arcs  $(k, l)$ , the matrix in branching vertex is reduced to matrix  $D$  after replacement  $d_{kl} \neq \infty$  on  $d_{kl} = \infty$ . If  $C^* > \varphi(\overline{k}, l)$  then in a searching tree a vertex  $\{(\overline{k}, l)\}$  is added to the branching vertex together with a reduced matrix  $D$  and a lower bound  $\varphi(\overline{k}, l)$  of cost of detours  $\{(\overline{k}, l)\}$ , otherwise it has no further prolongation. The bound  $\varphi(\overline{k}, l)$  is equal to the lower bound of vertex of branching increased by the sum of reduction factors.

8. For determination of the subset of detours  $\{(k, l) \notin R'$  which contain arc  $(k, l)$ , in a matrix at branching vertex the line  $k$  and a column  $l$  is excluded; lines  $s, k$  and columns  $k, l$ , if  $\{s, k\} \in R$ , are excluded; lines  $k, l$  and columns  $p, l$ , if  $\{l, p\} \in R$ , are excluded; lines  $s, k, l$  and columns  $k, l, p$ , if  $\{s, k\}, \{l, p\} \in R$ , are excluded. Matrix  $D$  is searched by reduction of the obtained matrix. After an evaluation of the lower bound  $\varphi((k, l) \notin R')$  of cost of detours  $\{(k, l) \notin R'\}$  which is defined under formulas (5), a vertex  $\{(k, l) \notin R'\}$  with the matrix  $D$  and a bound  $\varphi((k, l) \notin R')$  at performance of condition  $C^* > \varphi((k, l) \notin R')$  joins branching vertex, otherwise it is excluded from review.

9. If dimension of matrix  $D$  is equal to 1, admissible solution of HRPP is obtained. If the estimation of vertex of a branching tree is less than  $C(z(R))$ , it is remembered obtained admissible solution  $z(R)$ . We appropriate an estimation of current vertex as  $C(z(R))$ . We exclude from the further reviewing the vertices which have estimation more or equal  $C^*$ . If there is no remained vertex after elimination, go to step 10. Otherwise, it's necessary to find a vertex of searching tree with the minimal estimation. If there are such vertices, chose vertex which has the greatest depth.

10. If no admissible solution is found, then HRPP is unsolvable. Otherwise  $z^*(R)$  - a required Hamiltonian cycle, and  $C^*$  - its weight.

### Example

Graph  $H = (V, U)$  is defined by the matrix of costs

$$[d_{ij}]_6 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} \infty & 10 & \infty & 5 & 7 & 18 \\ 10 & \infty & 23 & 15 & \infty & 8 \\ \infty & 23 & \infty & 25 & 24 & \infty \\ 5 & 15 & 25 & \infty & \infty & 23 \\ 7 & \infty & 24 & \infty & \infty & 11 \\ 18 & 8 & \infty & 23 & 11 & \infty \end{bmatrix} \end{matrix}.$$

Subset of edges  $R = \{\{3, 4\}, \{1, 5\}, \{2, 6\}\}$ . It is required to find a solution of HRPP or to detect that it has no solution.

1. We reduce given matrix of costs and obtain reduced matrix

$$[d_{ij}]_6 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} \infty & 5 & \infty & 0 & 1 & 13 \\ 2 & \infty & 0 & 7 & \infty & 0 \\ \infty & 0 & \infty & 2 & 0 & \infty \\ 0 & 10 & 5 & \infty & \infty & 18 \\ 0 & \infty & 2 & \infty & \infty & 4 \\ 10 & 0 & \infty & 15 & 2 & \infty \end{bmatrix} \end{matrix} .$$

We calculate the lower bound of costs of all HRPP solutions:

$$\begin{aligned} \varphi(R) &= \min\{d_{34}, d_{43}\} + \min\{d_{15}, d_{51}\} + \min\{d_{26}, d_{62}\} + \sum_{i=1}^6 \alpha_i + \sum_{j=1}^6 \beta_j = \\ &= \min\{2, 5\} + \min\{1, 0\} + \min\{0, 0\} + 56 + 16 = 74 . \end{aligned}$$

$\Delta_{34} = \min\{d_{34}, d_{43}\} = d_{34} = 2 > 0$ , therefore let  $d_{34} = 0$ ,  $d_{43} = 3$ , we obtain completely reduced matrix

$$[d_{ij}]_6 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} \infty & 5 & \infty & 0 & 1 & 13 \\ 2 & \infty & 0 & 7 & \infty & 0 \\ \infty & 0 & \infty & 0 & 0 & \infty \\ 0 & 10 & 3 & \infty & \infty & 18 \\ 0 & \infty & 2 & \infty & \infty & 4 \\ 10 & 0 & \infty & 15 & 2 & \infty \end{bmatrix} \end{matrix} .$$

2. We calculate estimations for each zero elements of completely reduced matrix:

$$\gamma(1, 4) = 1, \gamma(2, 3) = 2, \gamma(2, 6) = 4, \gamma(3, 2) = 0, \gamma(3, 4) = 0, \gamma(3, 5) = 1, \gamma(4, 1) = 3, \gamma(5, 1) = 2, \gamma(6, 2) = 2 .$$

Arc  $(2, 6)$  of multigraph  $H'$  initiates branching.

3.  $\{2, 6\} \in R$ .

4.  $d_{26} \neq \infty$ . The set of all solutions is divided into the subset of detours  $\{(6, 2)\}$ , which contains arc  $(6, 2)$  and a subset of detours  $\{(2, 6)\}$ , which contains arc  $(2, 6)$ .

5. We calculate a matrix which defines a subset of detours  $\{(6, 2)\}$ . We exclude the arc  $(2, 6)$ , let  $d_{26} = \infty$ , also we exclude line 6 and column 2. We reduce the obtained matrix and obtain new matrix

$$[d_{ij}]_5 = \begin{matrix} & \begin{matrix} 1 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} \infty & \infty & 0 & 1 & 9 \\ 2 & 0 & 7 & \infty & \boxed{\infty} \\ \infty & \infty & 0 & 0 & \infty \\ 0 & 3 & \infty & \infty & 14 \\ 0 & 2 & \infty & \infty & 0 \end{bmatrix} \end{matrix} .$$

This matrix gives estimation

$$\varphi(6, 2) = \varphi(R) + d_{62} + \sum_{i \neq 6} \alpha'_i + \sum_{j \neq 2} \beta'_j = 74 + 0 + 4 = 78 .$$

6. We obtain a matrix which defines the subset of detours  $\{(2, 6)\}$ . To the arc  $(6, 2)$  we assign the weight  $d_{62} = \infty$ , exclude line and column 6. We reduce the obtained matrix:

$$[d_{ij}]_5 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} \infty & 5 & \infty & 0 & 1 \\ \infty & 0 & \infty & 0 & 0 \\ 0 & 10 & 1 & \infty & \infty \\ 0 & \infty & 0 & \infty & \infty \\ 8 & \boxed{\infty} & \infty & 13 & 0 \end{bmatrix} \end{matrix} .$$

We find an estimation:

$$\varphi(2,6) = \varphi(R) + d_{26} + \sum_{i \neq 2} \alpha_i' + \sum_{j \neq 6} \beta_j' = 74 + 0 + 2 + 2 = 78 .$$

9. Dimension of the reduced matrix is more than 1. Any of vertices  $\{(2,6)\}$ ,  $\{(6,2)\}$  of branching tree can be chosen as the active. We choose the vertex  $\{(2,6)\}$ .

2. We calculate estimations for each zero elements of the matrix:  $\gamma(1,4) = 1$ ,  $\gamma(3,2) = 5$ ,  $\gamma(3,4) = 0$ ,  $\gamma(3,5) = 0$ ,  $\gamma(4,1) = 1$ ,  $\gamma(5,1) = 0$ ,  $\gamma(5,3) = 1$ ,  $\gamma(6,5) = 8$ .

Arc  $(6,5)$  initiates branching.

3.  $\{6,5\} \notin R$ ; the set of detours  $\{(2,6)\}$  is divided into the subset of detours  $\{(6,5) \notin R'\}$  which includes arc  $(6,5)$ , and a subset of detours  $\{\overline{(6,5)}\}$  which does not contain this arc.

7. In the matrix which defines a subset of detours  $\{(2,6)\}$ , let  $d_{65} = \infty$  and reduce obtained matrix. Outcome of such operations is the matrix

$$[d_{ij}]_5 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} \infty & 5 & \infty & 0 & 1 \\ \infty & 0 & \infty & 0 & 0 \\ 0 & 10 & 1 & \infty & \infty \\ 0 & \infty & 0 & \infty & \infty \\ 0 & \infty & \infty & 5 & \boxed{\infty} \end{bmatrix} \end{matrix}$$

and estimation  $\varphi(\overline{(6,5)}) = \varphi(2,6) + S(6,5) = 78 + 8 = 86$  which limits from below costs of all detours  $\{\overline{(6,5)}\}$ .

Set  $\{\overline{(6,5)}\}$  includes arc  $(2,6)$  and does not contain arc  $(6,5)$ .

8. In a matrix which corresponds to the lower bound  $\varphi(2,6)$ , we exclude line 6 and column 5. As edges  $\{6,5\} \notin R$  and  $\{5,1\} \in R$  are adjacent, we exclude line 5 and column 1 and we forbid arc  $(1,2)$  assigning  $d_{12} = \infty$  to exclude subcycles. We obtain a matrix

$$[d_{ij}]_3 = \begin{matrix} & \begin{matrix} 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \boxed{\infty} & \infty & 0 \\ 0 & \infty & 0 \\ 9 & 0 & \infty \end{bmatrix} \end{matrix}$$

and estimation  $\varphi(\{(6,5) \notin R'\}) = \varphi(2,6) + d_{51} + \sum \alpha_i' + \sum \beta_j' = 78 + 0 + 1 = 79$ .

9. Dimension of this matrix is more than 1, we select the vertex  $\varphi(6,2)$  which has the minimal lower estimation.

2. We calculate an estimation for each zero element:



$$\gamma(1,4) = 1, \gamma(2,3) = 4, \gamma(3,4) = 0, \gamma(3,5) = 1, \gamma(4,1) = 5, \gamma(5,1) = 0, \gamma(5,6) = 9.$$

The further branching is fulfilled by means of an arc  $(5,6)$ .

3.  $\{5,6\} \notin R$ , the set of detours  $\{(6,2)\}$  is presented by a division into subset  $\{(5,6) \notin R'\}$  and subset  $\{(\overline{5,6})\}$ .

7. To vertex  $\{(\overline{5,6})\}$  corresponds the matrix

$$[d_{ij}]_5 = \begin{matrix} & \begin{matrix} 1 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} \infty & \infty & 0 & 1 & 9 \\ 2 & 0 & 7 & \infty & \infty \\ \infty & \infty & 0 & 0 & \infty \\ 0 & 3 & \infty & \infty & 14 \\ 0 & 2 & \infty & \infty & \boxed{\infty} \end{bmatrix} \end{matrix}$$

and an estimation  $\varphi(\overline{5,6}) = \varphi(6,2) + S(5,6) = 78 + 9 = 87$ .

8. In the matrix which describe set of detours  $\{(5,6) \notin R'\}$ , we assign  $d_{21} = \infty$  to exclude subcycles:

$$[d_{ij}]_3 = \begin{matrix} & \begin{matrix} 1 & 3 & 4 \end{matrix} \\ \begin{matrix} 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \boxed{\infty} & 0 & 7 \\ \infty & \infty & 0 \\ 0 & 3 & \infty \end{bmatrix} \end{matrix}.$$

We calculate an estimation of vertex:  $\varphi((5,6) \notin R') = \varphi(6,2) + d_{15} = 78 + 1 = 79$ .

9. For branching we choose vertex  $\{(6,5) \notin R'\}$  with the minimal lower bound.

2. We calculate estimations for each zero element of the matrix which corresponds to current vertex:

$$\gamma(1,4) = 0, \gamma(3,2) = 9, \gamma(3,4) = 0, \gamma(4,3) = 9.$$

Branching is continued by an arc  $(3,2) \notin R'$ .

7. Matrix

$$[d_{ij}]_3 = \begin{matrix} & \begin{matrix} 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \infty & \infty & 0 \\ \boxed{\infty} & \infty & 0 \\ 0 & 0 & \infty \end{bmatrix} \end{matrix}$$

defines set of detours  $\{(\overline{3,2})\}$  with lower bound  $\varphi(\overline{3,2}) = \varphi(6,5) + S(3,2) =$

$$= 79 + 9 = 88$$

8. In a matrix which corresponds to vertex  $\{(\overline{6,5}) \notin R'\}$ , we assign  $d_{32} = \infty$  and exclude lines 3, 4 and columns 2, 3.

9. As a result we obtain a matrix of the one element  $(1,4)$  and,  $\{(2,6)\}$ . Therefore we obtain detour  $(2,6), (6,5),$

$(5,1), (1,4), (4,3), (3,2)$  with cost  $\varphi((3,2) \notin R') = \varphi(\overline{6,5}) + d_{43} = 79 + 0 = 79$ .

All vertices of the branching tree have estimations, greater or equal 79, go to step 10.

10. The obtained admissible solution of problem

$z(R) = \{(2,6), (6,5), (5,1), (1,4), (4,3), (3,2)\}$  with cost  $C(z(R)) = 79$  is optimum.

The Fig. 1 illustrates the solution tree of this example.

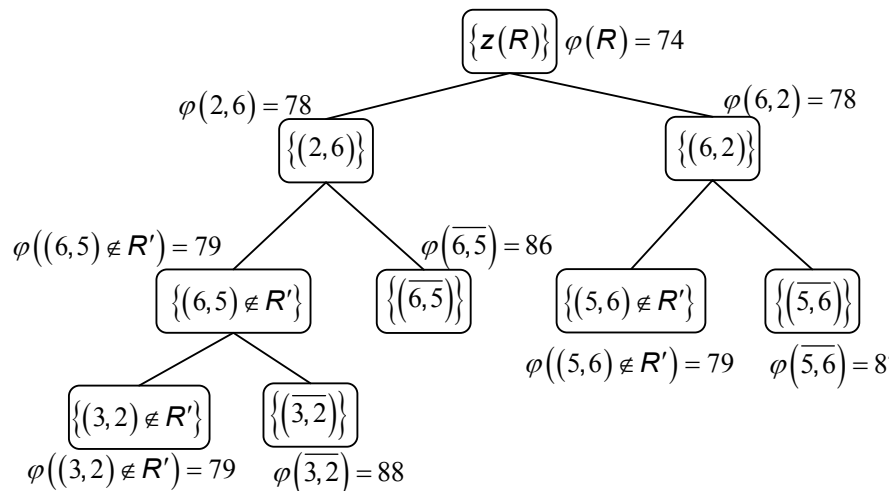


Fig. 3. The Solution Tree

## Conclusion

The method has been implemented in C# programming language. For performance tests we have used Intel Core 2 Duo 1,6GHz/2Gb RAM PC. Time of solving the Travelling Salesman Problem for dimension 50 occupies an average of 5 minutes.

## Bibliography

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. – М.: Мир, 1982. – 416с.
2. Garashchenko I., Panishev A.: Method of Finding Hamilton Routes in Transport Network // Artificial Intelligence and Decision Making. – ITHEA, Sofia, 2008. – Number 7. – P. 43–48
3. Кофман Е. Теория расписаний и вычислительные машины. – М.: Наука, 1984. – 334 с.
4. Гаращенко И.В., Морозов А.В., Панишев А.В. Метод решения гамильтоновой задачи коммивояжера // Искусственный интеллект. Выпуск 3. – Донецк: ИПШИ, 2008. – С. 630 - 637.
5. Яблонский А. А. Минимизация кольцевых маршрутов доставки продукции потребителям // Экономика и математические методы. – 2006. – Т. 42, № 3. – С. 124 - 131.

## Authors' Information



**Andriy Vasilyovich Morozov** – teacher-assistant of the Chair of informatics and computer modelling, aspirant, Zhytomyr State Technological University, Chernyahovskogo 103, Zhytomyr, Ukraine, 10005, e-mail: morozov.andriy@gmail.com



**Anatoliy Vasilyovich Panishev** – head of the Chair of informatics and computer modelling, professor, doctor of technical sciences, Zhytomyr State Technological University, Chernyahovskogo 103, Zhytomyr, Ukraine, 10005

## COMPARATIVE ANALYSIS OF STATISTICAL PROPERTIES OF THE HURST EXPONENT ESTIMATES OBTAINED BY DIFFERENT METHODS

Ludmila Kirichenko, Tamara Radivilova

**Abstract:** Estimating of the Hurst exponent for experimental data plays a very important role in research of processes which show properties of self-affinity. There are many methods for estimating the Hurst exponent using time series. The aim of this work is to carry out the comparative analysis of statistical properties of estimates of the Hurst exponent obtained by different methods using short length model fractal time series (the number of values less than 4000). In the work the most common used methods for estimating the Hurst exponent are researched. They are: *R/S*-analysis, variance-time analysis, detrended fluctuation analysis (DFA) and wavelet-based estimation. The fractal Brownian motion that is constructed using biorthogonal wavelets has been chosen as a model random process which exhibit fractal properties.

In the work the results of a numerical experiment are represented where the fractal Brown motion was modelled for the specified values of the exponent  $H$ . The values of Hurst exponent for the model realizations were varied within the whole interval of possible values  $0 < H < 1$ . The lengths of the realizations were defined as 500, 1000, 2000 and 4000 values. The estimates of  $H$  were calculated for each generated time series using the methods mentioned above. Samples of estimates of the exponent  $H$  were obtained for each value of  $H$  and their statistical characteristics were researched.

The results of the analysis have shown that the estimates of the Hurst exponent, which were obtained for the realisations of short length using the considered methods, are biased normal random variables. For each method the bias depends on the true value of degree of self-similarity of a process and a length of time series. Those estimates which are obtained by the DFA method and the wavelet transform have the minimal bias. Standard deviations of the estimates depend on the estimation method and decrease while the length of the series increases. Those estimates which are obtained by using the wavelet analysis have the minimal standard deviation.

**Keywords:** Hurst exponent, estimate of the Hurst exponent, self-similar stochastic process, time series, methods for estimating the Hurst exponent

**ACM Classification Keywords:** G.3 Probability and statistics - Time series analysis, Stochastic processes, G.1 Numerical analysis, G.1.2 Approximation - Wavelets and fractals

---

### Introduction

Nowadays problems of nonlinear physics, radio electronics, control theory, image processing require development and employment of new mathematical models, methods and algorithms for data analysis. At present time it has been generally accepted, that many of stochastic processes in nature and in engineering exhibit long-range dependence and fractal structure. The most suitable mathematical method for research of dynamics and structure of such processes is fractal analysis.

Stochastic process  $X(t)$  is statistically self-similar if the process  $a^{-H}X(at)$  show the same second-order statistical properties as  $X(t)$ . Long-range dependence means slow (hyperbolic) decay in time of the autocorrelation function of a process. The parameter  $H$  ( $0 < H < 1$ ) is called the Hurst exponent and is a measure of self-similarity or a measure of duration of long-range dependence of a stochastic process.

For values  $0.5 < H < 1$  time series demonstrates persistent behaviour. In other words, if the time series increases (decreases) in a prior period of time, then this trend will be continued for the same time in future with the probability which increases the more the Hurst exponent gets greater than 0.5. The value  $H = 0.5$  indicates the independence (the absence of any memory about the past) of values of time series. The interval  $0 < H < 0.5$  corresponds to antipersistent time series: if a system demonstrates growth in a prior period of time,

then it is likely to fall in the next period with the probability which increases the less the Hurst exponent gets below 0.5.

It is obvious that estimation of the Hurst exponent for experimental data plays an important role in study of processes which exhibit properties of self-similarity. There are many methods for evaluation of the Hurst exponent for time series. Sufficient review of these methods is represented in [Willinger, 1996; Clegg, 2005]. Nevertheless, at present time there is no a proper summary research where the results of the estimation of the Hurst exponent  $H$  with different methods would be generalized and the comparative analysis of statistical properties of estimations obtained for small amount of sample data would be. The given work is an attempt to carry out such research for the most common used methods of estimation of self-similarity.

The aim of this work is to carry out the comparative analysis of statistical properties of estimates of the Hurst exponent obtained by different methods using short length model fractal time series (the number of values less than 4000). In the work the most common used methods for estimating the Hurst exponent are researched. They are:  $R/S$ -analysis (rescaled range method) (see, for instance [Feder, 1988; Peters, 1996; Stollings, 2003; Sheluhin, 2007], variation in time of variance of an aggregate time series (variance-time analysis), see [Stollings, 2003; Sheluhin, 2007], detrended fluctuation analysis (see [Kantelhardt, 2001; Chen, 2002; Gu, 2006; Kantelhardt, 2008]) and estimation using the wavelet analysis (see [Mallat, 1998; Abry, 1998; Abry, 2003]) The fractal Brownian motion has been chosen as a model random process which exhibit fractal properties.

### Methods of estimating the Hurst exponent

**Rescaled range method.** According to this method for the time series  $x(t)$  of the length  $\tau$  the rate  $\frac{R(\tau)}{S(\tau)}$  is defined, where  $R(\tau)$  is the range of the cumulative deviate series  $x^{cum}(t, \tau)$ ,  $S(\tau)$  is standard deviation of the initial series:

$$R/S = \frac{\max(x^{cum}(t, \tau)) - \min(x^{cum}(t, \tau))}{\sqrt{\frac{1}{\tau-1} \sum_{t=1}^{\tau} (x(t) - \bar{x})^2}}, \quad t = \overline{1, \tau}, \quad (1)$$

where  $\bar{x}(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} x(t)$ ,  $x^{cum}(t, \tau) = \sum_{i=1}^t x(i) - \bar{x}(\tau)$ .

For a self-similar process and big values of  $\tau$  this ratio has the following characteristics:

$$M \left[ \frac{R}{S} \right] \sim (c \cdot \tau)^H, \quad (2)$$

where  $c$  is a constant.

The log-log diagram of dependence of  $\frac{R(\tau)}{S(\tau)}$  on  $\tau$  represents a line approximated by the least square method.

Then the estimate of the exponent  $H$  is a tangent of angle of slope of the line which represents the dependence of  $\log \frac{R(\tau)}{S(\tau)}$  on  $\log(\tau)$ .

**Variance-time analysis.** When someone mentions the aggregation of time series  $x(t)$  of the length  $\tau$  on the time scale with the parameter  $m$ , they mean the transition to the process  $x^{(m)}$ , where  $x_k^{(m)} = \frac{1}{m} \sum_{t=km-m+1}^{km} x(t)$ ,  $k = \overline{1, \tau/m}$ . For self-similar process the variance of the aggregated time series  $x^{(m)}$  for big values of  $m$  follows the formula:

$$\text{Var}(x^{(m)}) \sim \frac{\text{Var}(x)}{m^\beta}. \quad (3)$$

In this case the parameter of self-similarity  $H = 1 - \frac{\beta}{2}$  can be obtained if to generate an aggregated process on different levels of aggregation  $m$  and calculate the variance for each level. The dependence diagram of  $\log(\text{Var}(x^{(m)}))$  on  $\log(m)$  will represent a line with a slope equal to  $-\beta$ .

**Detrended fluctuation analysis (DFA).** According to the DFA method, for the initial time series  $x(t)$  the cumulative time series  $y(t) = \sum_{i=1}^t x(i)$  is constructed which is then divided into  $N$  segments of length  $\tau$ , and for each segment  $y(t)$  the following fluctuation function is calculated:

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (y(t) - Y_m(t))^2, \quad (4)$$

where  $Y_m(t)$  is a local  $m$ -polynomial trend within the given segment. The averaged on the whole of the time series  $y(t)$  function  $F(\tau)$  depends on the length of the segment:  $F(\tau) \propto \tau^H$ .

In some interval the diagram of dependence of  $\log F(\tau)$  on  $\log \tau$  represents a line approximated by the least square method. Estimate of the exponent  $H$  is a tangent of the angle of slope of the line which represents the dependence of  $\log F(\tau)$  on  $\log(\tau)$ .

**Wavelet-based estimation.** According to the discrete wavelet transform, the time series  $x(t)$ ,  $(t=1, 2, \dots, n)$  consists of the collection of approximation and detail coefficients. The detail wavelet coefficients  $d_x(j, k)$  on the scale level  $j$  are defined in the following way:

$$d_x(j, k) = 2^{j/2} \sum_{i=1}^{n_j} x(i) \Psi_0(2^{-j}n - k), \quad (5)$$

where  $n_j = 2^{-j}n$  is a number of the wavelet coefficients on the level  $j$ ,  $\Psi_0$  is base wavelet function.

For estimation of the Hurst exponent in applied research the method described in [Abry, 1998] is the most used. The mentioned method is based on the statement that the averaged squared values of the wavelet coefficients

$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_x(j, k)|^2$  obey the scaling law:

$$\mu_j \sim 2^{(2H-1)j}, \quad (6)$$

where  $H$  is the Hurst exponent. The following equation represents the practicable method of the estimation of the Hurst exponent:

$$\log_2 \mu_j = \log_2 \left( \frac{1}{n_j} \sum_{k=1}^{n_j} |d_x(j, k)|^2 \right) \sim (2H - 1)j + \text{const}. \quad (7)$$

From this formula it can be concluded that if there is the long-range dependence of the time series  $x(t)$  then the Hurst exponent  $H$  can be obtained by estimating the slope of the graph of the function  $\log_2(\mu_j)$  from  $j$ .

---

### Modelling of fractal Brownian motion

One of the most known and simple models of stochastic dynamics which exhibits fractal properties is fractal Brownian motion (fBm). It is widely used in physics, chemistry, biology, economics and theory of network traffic.

Gaussian process  $X(t)$  is called fractal Brownian motion with the parameter  $H$ ,  $0 < H < 1$ , if the increments of

the random process  $\Delta X(\tau) = X(t + \tau) - X(t)$  are distributed in the following way

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi\sigma_0^2\tau^{2H}}} \cdot \int_{-\infty}^x \text{Exp}\left[-\frac{z^2}{2\sigma_0^2\tau^{2H}}\right] dz, \quad (8)$$

where  $\sigma_0$  is diffusion coefficient.

fBm with the parameter  $H = 0.5$  coincides with the classic Brownian motion. Increments of fBm are called fractal Gaussian noise and its dispersion can be described by the formula  $D[X(t + \tau) - X(t)] = \sigma_0^2 \tau^{2H}$ .

There are many methods of construction of fBm for the case of discrete time, which have been considered in [Mandelbrot, 1983; Feder, 1988; Voss, 1988; Cronover, 2000]. These models have some weak sides. One of them is underestimating (overestimating) of the degree of self-similarity of a process for small and big theoretical values of the Hurst exponent and short length of a model realisation [Jeongy, 1998; Cronover, 2000; Sheluhin, 2007].

One of the methods which can help to resolve the mentioned problems is construction of fBm using biorthogonal wavelets [Sellan, 1995; Abry, 1996; Sellan, 1995; Meyer, 1999; Bardet, 2003]. In this case the fBm realization is constructed using discrete wavelet transform where the detail wavelet coefficients on each level are independent normal random values and approximation wavelet coefficients are obtained using fractal autoregression and moving average process FARIMA:

$$B_H(t) = \sum_{k=-\infty}^{\infty} \Phi_H(t-k) S_k^{(H)} + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} 2^{-jH} \Psi_H(2^j t - k) \varepsilon_{j,k} - b_0, \quad (9)$$

where  $\Psi_H$  is biorthogonal base wavelet function,  $\Phi_H$  is corresponding  $\Psi_H$  scaling function,  $S_k^{(H)}$  is stationary Gaussian process FARIMA with the fractal differentiation parameter  $d = H - 0.5$ ,  $\varepsilon_{j,k}$  – independent standard Gaussian random values,  $b_0$  – constant where  $B_H(0) = 0$ .

---

## Results of the investigation

---

In the work the results of numerical experiment are represented where fractal Brownian motion with the specified value of exponent  $H$  have been simulated. The values of  $H$  for model realisation were varied within the whole interval  $0 < H < 1$ . Length of realizations was accepted equal to 500, 1000, 2000 and 4000 values. For each received realization estimates  $H$  have been obtained using the methods described above:  $R/S$ -analysis ( $\hat{H}_{rs}$ ), variance-time analysis ( $\hat{H}_d$ ), DFA method ( $\hat{H}_{fa}$ ) and discrete wavelet transform ( $\hat{H}_w$ ). For each value of  $H$  samples of its estimates have been computed, and their statistical characteristics have been investigated.

### Investigation of bias of estimates

Fig. 1 shows dependence of average values of estimates of the Hurst exponent on its theoretical value. Model realizations contained 1000 values. Solid line refers to the theoretical values of  $H$ . Obviously, average values of the estimates are biased, where the bias depends on the theoretical values of the Hurst exponent.

Obviously, the estimates of the Hurst exponent are biased in a region of persistence as well as in a antipersistence one. Since most of fractal processes have the long-range dependence, we will be considering results only for the interval  $0.5 < H < 1$ . From Fig. 1 we can see that the estimates obtained by the methods of  $R/S$ -analysis and variance-time analysis are the most biased.

Let us consider the results of estimation of the exponent  $H$  by the method of  $R/S$ -analysis. The method of rescaled range proposed by Hurst is, perhaps, the most popular one and used in all fields of scientific research. Its main merit is its robustness. Actually, this method works even on non-stationary data. But also, as it was noticed by Hurst, the estimates of  $H$  below  $H \approx 0.75$  obtained by  $R/S$ -method are overestimated, and the estimate of  $H$  over  $H \approx 0.75$  are understate.

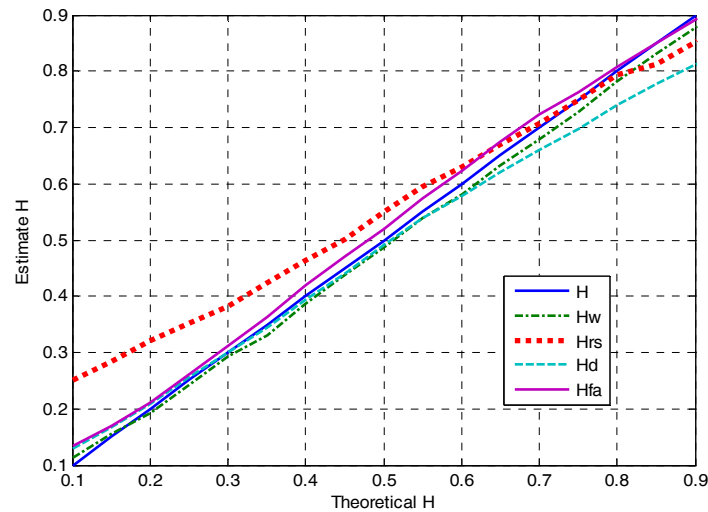


Figure 1. Dependence of average values of estimates obtained by various methods on theoretical values of  $H$

Fig. 2 represents dependence of average values of estimates  $\hat{H}_{rs}$  on theoretical values of  $H$  for model series of different length. Obviously, average values of estimates can be approximates quite well with lines  $\hat{H}_N = k_N H + b_N$ , where coefficients  $k_N$  and  $b_N$  depend on realization  $N$  where the estimation is done. This lines cross the line of theoretical values of  $H$  at around  $H \approx 0,75$ ; and are overestimated below this values and are underestimated over this value. The results of the performed research confirm the results obtained by analysing the estimation another models [Feder, 1988; Jeongy, 1998; Кириченко, 2005; Sheluhin, 2007]. With the increase of the realisation length  $N$  the angle of slope  $k_N$  of the approximated line increases slowly and approaches the theoretical value  $\pi/4$ .

Due to its simplicity and easy understanding of its results the method of variance-time analysis is the most common used for assessment of self-similarity of the information network traffic. Nevertheless, for processes with long-range dependence this method gives undervalued estimates. [Jeongy, 1998; Кириченко, 2005; Sheluhin, 2007]. This can be unacceptable, for instance, in case of assessment of the network load during the transmission of the self-similar traffic. [Stollings, 2003].

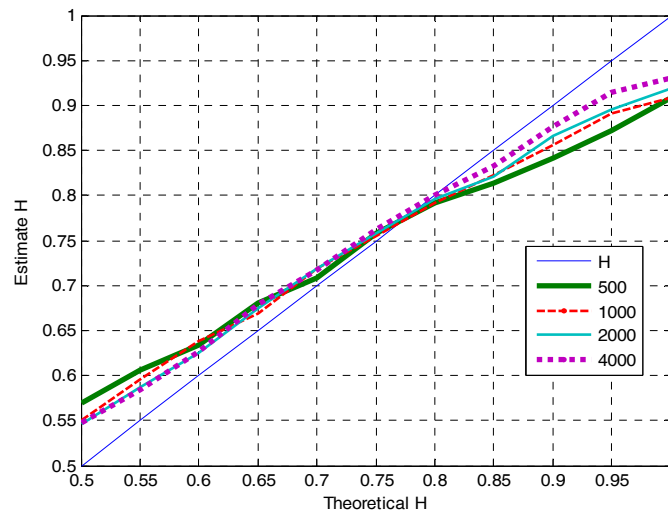


Figure 2. Dependence of average values of estimates  $\hat{H}_{rs}$  on theoretical value of  $H$

Fig. 3 shows the dependence of the average values of estimates  $\hat{H}_d$  obtained by the method of variance-time analysis. These dependence can also be approximated by the lines  $\hat{H}_N = k_N H + b_N$  where coefficients  $k_N$  and  $b_N$  depend on the length of realization  $N$ . In this case the approximating lines cross the line of theoretical values at around  $H \approx 0,5$  (see Fig. 1) and actually the estimates of the exponent  $H$  are undervalued within the whole interval of persistence. The bias increases with the growth of the Hurst exponent, particularly for  $H > 0,9$ . It can be noted that the bias of the estimates  $\hat{H}_d$  are greater than the appropriate bias of  $\hat{H}_{rs}$ . With the increase of the realization length  $N$  the bias slowly decreases.

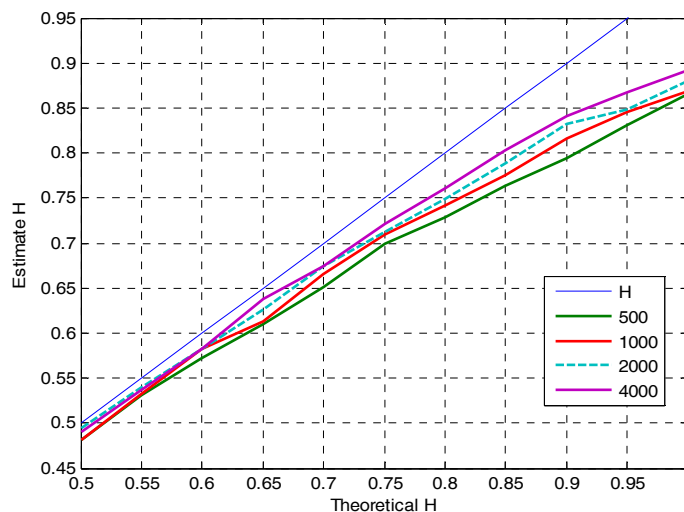


Figure 3. Dependence of average values of estimates  $\hat{H}_d$  on theoretical value of  $H$

Method DFA is based on the ideology of one-dimensional random walk and is widely used in analysis of bioelectric signals. Estimates  $\hat{H}_{fa}$  obtained by DFA method can be characterised by small bias within the interval  $0,5 < H < 0,9$  (see Fig. 4) even for realizations of short length. The sign of this bias reverses and increases for  $H > 0,9$ . It should be brought into focus that the most of natural and information fractal processes have the degree of self-similarity less than 0,9.

Methods of estimation of the Hurst exponent by the use of wavelet analysis are the most recent and still have not been commonly used. Nevertheless, their merits are obvious. In the work [Abry, 1998] it has been shown that the estimates are asymptotically unbiased if base wavelets are chosen in a proper way. Fig 5 represents dependence of the average values of  $\hat{H}_w$  obtained by the method of discrete wavelet expansion with the base wavelet function of Daubechies D4. It is obvious that with the increase of the time series length  $N$  the bias decreases and is actually equal to 0 for  $N \approx 4000$ .



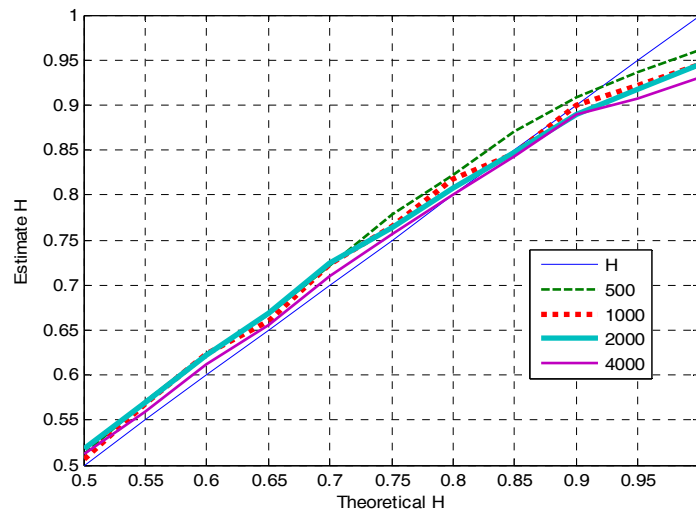


Figure 4. Dependence of the average values of the estimates  $\hat{H}_{fa}$  on the theoretical value of  $H$

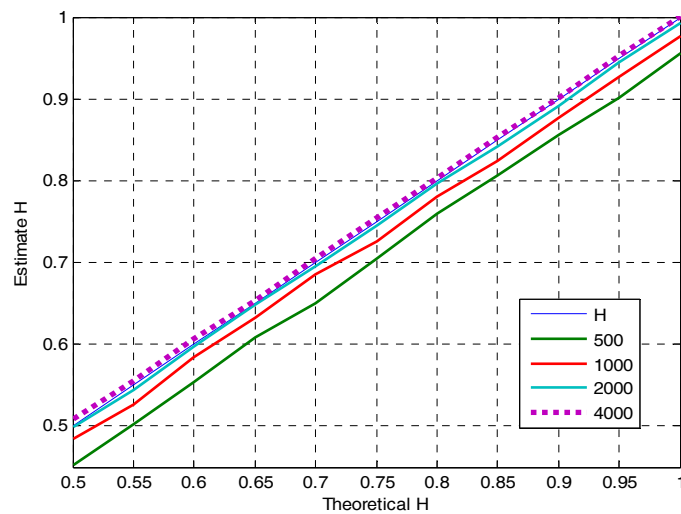


Figure 5. Dependence of the average values of the estimates  $\hat{H}_w$  on the theoretical values of  $H$

**Research of standard deviations of the estimates**

In the work the dependence of standard deviations of estimates of the Hurst exponent on the values of  $H$  and length of the model fractal series has been investigated for each method. In Table 1 the values of standard deviations of the estimates of the Hurst exponent which have been received for the series of length of 1000 values are represented.

Table 1. Standard deviations of the estimates of the Hurst exponent

Estimation method	Range $S_{\hat{H}}$	Dependence on $H$
$R/S$ -analysis	$0.03 \leq S_{\hat{H}} \leq 0.08$	Increases along with $H$
Variation of dispersion	$S_{\hat{H}} \approx 0.06$	No obvious trend
DFA	$S_{\hat{H}} \approx 0.07$	No obvious trend
Wavelet analysis	$S_{\hat{H}} \approx 0.045$	No obvious trend

Table 2 demonstrates how standard deviations obtained during the estimation of  $\hat{H}$  (rounded value) decrease with the increase of the length of time series. In this specific case the model Hurst exponent  $H = 0.8$ .

Table 2. Standard deviations of the estimates depending on length of time series

$S_{\hat{H}}$	500	1000	2000	4000
$S_{Hrs}$	0.08	0.06	0.05	0.04
$S_{Hd}$	0.07	0.06	0.05	0.045
$S_{Hfa}$	0.085	0.07	0.055	0.045
$S_{Hw}$	0.065	0.045	0.03	0.02

The problem of distribution law of the estimates of  $H$  was considered in number of works where it was shown, numerically and analytically, that the estimates are normal for a specific method or specific values of the Hurst exponent (see, for instance [Feder, 1988, Peters, 1996; Abry, 1998]). In this work the distribution of the estimates  $\hat{H}$  have been researched for each method and different values of the parameter. For all considered methods the hypothesis of normal distribution of sample values of the estimates with parameters  $N(\bar{H}, S_{\hat{H}})$  have been suggested. Nearly for all sample data the hypothesis has been accepted with the confidence level  $\alpha = 0.05$  by few fitting criterions.

## Conclusion

Thus, the estimates of the Hurst exponent which are obtained by the methods considered above are biased normal random variables. For each method the bias depends on a true value of degree of self-similarity and the length of a time series. Standard deviations of the estimates depend on the estimation method and decrease with the growth of the series length. Summarizing the results of the numerical research we can make a conclusion that the estimates with the least bias and standard deviation can be given by the method which uses wavelet analysis. Also, other methods have some merits which can be significant in relation to some aims and ways of the research. For instance,  $R/S$ -analysis allows to estimate the degree of self-similarity of a time series for which the wavelet estimation nearly cannot be applied, DFA method gives the best results for short series. Thus, in the most cases for the estimation of the Hurst exponent it makes sense to use various methods and comparison of the results provides an extra information.

## Bibliography

- [Abry, 1996] P. Abry. The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer. Remarks and fast implementation. P. Abry, F. Sellan. Applications and Computing Harmonic Analyse (V.3(4)), 1996.
- [Abry, 1998] P. Abry. Wavelet analysis of long-range dependent traffic. P. Abry, D. Veitch. IEEE/ACM Transactions Information Theory (№ 1(44)), 1998.
- [Abry, 2003] P. Abry. Self-similarity and long-range dependence through the wavelet lens. P. Abry, P. Flandrin, M.S. Taqqu, D. Veitch. Theory and applications of long-range dependence, Birkhäuser, 2003.
- [Bardet, 2003] J.-M. Bardet. Generators of long-range dependence processes: a survey, Theory and applications of long-range dependence. Bardet J.-M., G. Lang, G. Oppenheim, A. Philippe, S. Stoev, M.S. Taqqu. Theory and applications of long-range dependence, Birkhäuser, 2003.
- [Chen, 2002] Z. Chen. Effect of non-stationarities on detrended fluctuation analysis. Z. Chen, P.Ch. Ivanov, K. Hu, H.E. Stanley. Phys. Rev. E 65, 041107, 2002.

- [Clegg, 2005] R.G. Clegg. A practical guide to measuring the Hurst parameter. R. G. Clegg. Computing science technical report (№ CS–TR–916), 2005.
- [Cronover, 2000] R. M. Cronover. Introduction to fractals and chaos. R. M. Cronover. John and Bartlett Publishers, N.Y., 2000.
- [Feder, 1988] J. Feder. Fractals. J. Feder. Plenum, New York, 1988.
- [Gu, 2006] G.-F. Gu. Detrended fluctuation analysis for fractals and multifractals in higher dimensions. G.-F. Gu, W.-X. Zhou. Phys. Rev. (E 74, 061104), 2006.
- [Jeongy, 1998] H.-D. J. Jeongy. A Comparative Study of Generators of Synthetic Self-Similar Teletraffic. H.-D. J. Jeongy, McNickle D., Pawlikowski K. Department of Computer Science and Management, University of Canterbury, 1998.
- [Kantelhardt, 2001] J.W. Kantelhardt. Detecting long-range correlations with detrended fluctuation analysis. J.W. Kantelhardt, E. Koscielny-Bunde, H.H.A. Rego, S. Havlin, A. Bunde. Phys (A 295, 441), 2001.
- [Kantelhardt, 2008] J. W. Kantelhardt. Fractal and Multifractal Time Series. J. W. Kantelhardt. <http://arxiv.org/abs/0804.0747>, 2008.
- [Mallat, 1998] S. Mallat. A wavelet tour of signal processing. S. Mallat. Academic Press, San Diego, London, Boston, N.Y., Sydney, Tokyo, Toronto, 1998.
- [Mandelbrot, 1983] B. B. Mandelbrot. The Fractal Geometry of Nature. B. B. Mandelbrot. W.H. Freeman, San Francisco, 3rd Edition, 1983.
- [Meyer, 1999] Y. Meyer. Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion. Y. Meyer, F. Sellan, M.S. Taqqu. The Journal of Fourier Analysis and Applications 5(5), Birkhäuser, Boston 1999.
- [Peters, 1996] Edgar E. Peters. Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility. Edgar E. Peters. Wiley, 2 edition, 1996.
- [Sellan, 1995] F. Sellan. Synthèse de mouvements browniens fractionnaires par l'aide de la transformation par ondelettes. F. Sellan. Comptes Rendus de l'Academie des Sciences de Paris, Serie I, 1995.
- [Sheluhin, 2007] Oleg I. Sheluhin. Self-similar processes in telecommunications. Oleg I. Sheluhin, Sergey M. Smolskiy, Andrey V. Osin. JohnWiley & Sons Ltd, Chichester, 2007.
- [Stollings, 2003] W. Stollings. High-speed networks and Internets. Performance and quality of service. W. Stollings. New Jersey, 2002.
- [Voss, 1988] R. F Voss. Fractals in nature: From characterization to simulation in The Science of Fractal Images. R. F Voss. Springer-Verlag, New York, 1988.
- [Willinger, 1996] W. Willinger. Bibliographical guide to self-similar traffic and performance modeling for modern high-speed network in «Stochastic networks: theory and applications». W. Willinger, M. S. Taqqu, A. A. Erramilli. Claredon Press (Oxford University Press), Oxford, 1996.
- [Кириченко, 2005] Л. О Кириченко. Сравнительный анализ методов оценки параметра Херста самоподобных процессов Л. О. Кириченко, Т. А. Радивилова, М. И. Синельникова. Системи обробки інформації Вип. 8(48), Харьков, 2005.

---

## Authors' Information

---



**Ludmila Kirichenko** – Ph. D., Associate professor, Kharkiv National University of Radio Electronics; 14 Lenin Ave., 61166 Kharkiv, Ukraine; e-mail: [kirichenko.ludmila@gmail.com](mailto:kirichenko.ludmila@gmail.com).

*Major Fields of Scientific Research: Time series analysis, Stochastic self-similar processes, Wavelets and fractals, Chaotic systems*



**Tamara Radivilova** – Ph. D., lecturer, Kharkiv National University of Radio Electronics; 14 Lenin Ave., 61166 Kharkiv, Ukraine; e-mail: [tomachka\\_7@yahoo.com](mailto:tomachka_7@yahoo.com).

*Major Fields of Scientific Research: Computer-communication networks, Fractal processes, Time series analysis*

## SOME PROPERTIES OF ACYCLIC COMPOSITIONAL PROGRAMS

Tetiana Parfirova, Vadim Vinnik

**Abstract:** Structure, behaviour and properties of acyclic programs are investigated in methodological and mathematical framework of compositional programming. Relations of acyclic programs to sequential-parallel and chain programs (defined in the previous authors' papers) are investigated: any acyclic program can be transformed to an equivalent sequential-parallel program with extra usages of basic subroutines but without extra names (informally, enlarging the executable code but not enlarging data); acyclic program can be transformed to an equivalent chain program with extra names but without extra basic subroutine usages (informally, this enlarges data but does not enlarge the executable code); sequential composition of two acyclic programs can be transformed to an equivalent acyclic program with either complex internal interconnections (though without extra subroutines) or with one additional subroutine (though with simple interconnections).

**Keywords:** sequential composition, parallel composition, compositional programming, nominal function, acyclic program, sequential and parallel compositions.

**ACM Classification Keywords:** D.1.4 Sequential Programming, D.2.4 Software/Program Verification, F.3.2 Semantics of Programming Languages.

---

### Introduction

Previous papers [Parfirova 2010a, Parfirova 2010b] cover structure, properties and behavior of a certain simple form of programs where subroutines work sequentially, data flow in one direction from predecessors to successors, without loopback. It was noticed that this principle is intrinsic to the large-scale structural level of information processing systems, in particular to software systems that support business activity and electronic document processing, organizational activity, education etc. Except this, it was shown [Parfirova 2009] that sequential type of structure and its corresponding principle of operation is the most transparent implementation of the reference model of interaction in terms of entity platform [Redko 2008a, Redko *et al.* 2008]. Two particular cases, sequential-parallel and chain systems were described and investigated in [Parfirova 2010b]; it was shown that systems of the first kind can be transformed to the second kind, and the transformation doesn't introduce extra usages of subroutines to the compositional term that represents the system but requires extra names (used for temporary copies of data). The general model of acyclic programs is described in [Parfirova and Vinnik 2010]. The goal of this paper is to reveal some properties of acyclic programs and their relations to the previously investigated classes.

---

### Basic Notions: Compositional Programming

According to compositional programming (CP) and entity platform created on its basis [Redko 1978, Redko 1998, Redko 2008, Redko *et al.* 2008], data objects are modelled by means of nominal sets (NS), whereas programs and their parts as data transformers are modelled as nominal functions (NF).

Sets of all *names* and all *detonates* are designated as  $\mathbf{V}$  and  $\mathbf{D}$  respectively.  $V$ -NS is a finite functional<sup>1</sup> binary relation  $\alpha \in \mathbf{D}^V$  where  $V \subseteq \mathbf{V}$ . Note that, when  $\alpha$  is a  $V$ -NS,  $v \in V$  is some name,  $\alpha(v)$  is the value of that name in the NS. Let us denote the set of all NSs as  $\mathbf{N}$ .

By definition, NF is a unary partial function of the form  $\mathbf{N} \rightharpoonup \mathbf{N}$ . NF is called  $V$ -ary if  $\text{dom } f \subseteq \mathbf{D}^V$ , and  $(V, W)$ -ary if, except this,  $\text{rang } f \subseteq \mathbf{D}^W$ , for some  $V, W \subseteq \mathbf{V}$ . Not all NFs have an arity. Function that has some arity is called polyary. Here and further we investigate polyary NFs only.

---

<sup>1</sup>There is also a branch of CP called *compositional-nominative approach* where functionality of NS is not required [Nikitchenko, 1999; Nikitchenko, 2009].

Restriction is a  $(V, U \cap V)$ -ary NF  $\uparrow_U^V$  (where  $U, V \subseteq \mathbf{V}$ , vertical bar means restriction of a relation by its 1<sup>st</sup> component on a given set):

$$\uparrow_U^V(\alpha) = \alpha \mid U = \{(u, d) \mid (u, d) \in \alpha \wedge u \in U\}, \alpha \in \mathbf{D}^V.$$

For some mapping  $\xi: U \rightarrow V$ , renaming is a  $(V, U)$ -ary NF  $\downarrow^\xi$  (where  $U, V \subseteq \mathbf{V}$ ):

$$\downarrow^\xi(\alpha) = \{(u, \alpha(\xi(u))) \mid u \in U\}, \alpha \in \mathbf{D}^V.$$

Substantially, this operation replaces names  $v \in V$  in NS  $\alpha$  with new names  $u \in U$  such that  $\xi(u) = v$ . It should be noted that  $\xi$  does not have to be a one-to-one mapping. It means that one name from  $\alpha$  can turn after renaming into several new names. From the definition, it directly follows that if  $\text{pr}_1 \alpha = V \subseteq \text{pr}_2 \xi = W$  (where  $\text{pr}_i \rho$  means projection of a relation  $\rho$  by its  $i$ -th component) then  $(\uparrow_W^V \circ \downarrow^\xi)(\alpha) = \xi \circ \alpha$  where  $\circ$  means multiplication of unary functions (taking into account that NS  $\alpha$  can be regarded as a function that maps names to their denotates).

The identical mapping of some set  $X$  into itself is designated as  $\text{id}_X = \{(x, x) \mid x \in X\}$ .

Overlapping is a binary operation such that, for any NSs  $\alpha$  and  $\beta$ ,

$$\alpha \nabla \beta = \beta \cup \{(u, d) \mid (u, d) \in \alpha, u \notin \text{pr}_1(\beta)\}.$$

Operations on nominal functions are called *compositions*. Substantially, compositions model composing a complex program from simpler programs, or subroutines. Two fundamental compositions will be important in the next sections: multiplication (model of sequential execution) and overlapping (models parallel execution) defined as follows:

$$\begin{aligned} (f \circ g)(\alpha) &\cong g(f(\alpha)), \\ (f \nabla g)(\alpha) &\cong f(\alpha) \nabla g(\alpha). \end{aligned}$$

In the second formula,  $\nabla$  in the left side denotes a composition (an operation on nominal functions) whereas in the right side it means operation on nominal sets; this should not confuse the reader because the context of this operation is always clear. Let NFs  $f$  and  $g$  be  $(U_1, V_1)$ -ary and  $(U_2, V_2)$ -ary respectively. Then  $f \circ g$  is a  $(U_1, V_2)$ -ary NF if  $V_1 = U_2$  and totally undefined NF otherwise;  $f \nabla g$  is  $(U, V_1 \cup V_2)$ -ary NF if  $U_1 = U_2 = U$  and totally undefined otherwise.

---

### Sequential-Parallel and Chain Functions

---

Notions of sequential-parallel and chain functions were first introduced in [Parfirova 2010a, Parfirova 2010b]. Let us briefly remind the main definitions and theorems that will be important in this work.

Let  $\Phi$  be some set of NFs. NF  $h$  is called *sequential-parallel* in basis  $\Phi$  if it can be obtained from functions of the set  $\Phi$  and functions  $\uparrow_U^V, \downarrow^\xi, \text{id}_V$  by means of compositions  $\circ$  and  $\nabla$ .

*Link* with the kernel  $f$  is NF of the form

$$h = L_\zeta^\xi(f) = \downarrow^\zeta \nabla \left( \uparrow_{\text{pr}_2(\zeta)}^{\text{pr}_2(\xi)} \circ \downarrow^\xi \circ f \right),$$

where  $\xi, \zeta$  are renaming mappings,  $\xi: U \rightarrow V, \zeta: U' \rightarrow V'$ , where  $U, V, U', V' \subseteq \mathbf{V}$ . Instantially, the definition of the link means that some part of the input data, after selecting an appropriate part and renaming  $\zeta$ , is passed to the kernel function  $f$ , and the rest of input data is passed to the output through a renaming  $\xi$ . The overlapping of these two parallel branches gives the link's output. Note that renaming  $\zeta$  is applied to the whole input NS without excluding any names; except this, renaming can transform any input name to two or more output names. This means that renaming  $\zeta$  can be used to create "backup copies" of the input data, see fig. 1.

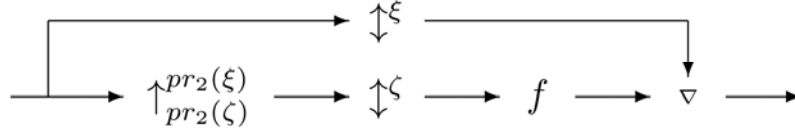


Fig. 1. Link

*Chain function* in basis  $\Phi$  (where  $\Phi$  is a set of NFs) is a function of the form  $h = h_1 \circ h_2 \circ \dots \circ h_n$  where every  $h_i$  ( $i = 1, \dots, n$ ) is a link with some kernel  $f \in \Phi$ .

It is proved [Parfirova 2010a, Parfirova 2010b] that any, say  $(U, V)$ -ary, sequential-parallel NF  $f$  in basis  $\Phi$  can be transformed into a chain function  $\hat{f}$  in the same basis with some arity  $(U, \hat{V})$  where  $V \subseteq \hat{V}$ , that is equivalent to  $f$  up to extra names, i.e.  $f = \hat{f} \circ \uparrow_V^{\hat{V}}$ .

### Acyclic Functions

Let us remind definitions [Parfirova and Vinnik 2010]. For an acyclic program consisting of  $n$  subroutines, let us consider the input of the whole program as the output of its imaginary 0-th subroutine; in a dual way, the output of the whole program is considered as the input of its imaginary  $(n+1)$ -th subroutine. The determinative feature of acyclic programs is that relation *act after* (or *use outputs of*) is a partial ordering on the set of all its subroutines. This order can be considered as linear: for each  $i, j$  where  $0 \leq i < j \leq n+1$ , data transfer channel from the output of the  $i$ -th subroutine to the input of the  $j$ -th subroutine exists if and only if  $i < j$ .

Suppose we have  $(U_i, V_i)$ -ary NFs,  $i = \overline{1, n}$ , a set of names  $V$  that will be also designated  $V_0$ , a set of names  $U$  that will be also designated  $U_{n+1}$ , and renaming mappings  $\xi_{ij}$  for all  $i, j$  where  $0 \leq i < j \leq n+1$ . The mappings  $\xi_{ij}$  can be regarded as an upper-right triangle matrix  $\Xi$  (this matrix has  $n+1$  rows and as many columns; row indices start with 0, and column indices start with 1). The following restrictions are imposed:

$$V_i \supseteq \bigcup_{j=i+1}^{n+1} \text{pr}_2(\xi_{ij}), \quad i = \overline{0, n}.$$

$$U_j = \bigcup_{i=0}^{j-1} \text{pr}_1(\xi_{ij}), \quad j = \overline{1, n+1}.$$

By definition, *acyclic NF* is NF  $f = T_{V,U}^{\Xi}(f_1, \dots, f_n)$  such that for any  $V$ -ary NS  $\alpha$  its value  $\beta = f(\alpha)$  is defined by the following equations (fig. 2) where  $\alpha_0 = \alpha$ ,  $\beta = \beta_{n+1}$ :

$$\beta_k = \nabla_{i=0}^{k-1} (\xi_{ik} \circ \alpha_i), \quad \text{for } k = \overline{1, n+1},$$

$$\alpha_k = f_k(\beta_k), \quad \text{for } k = \overline{1, n}.$$

Note that  $\alpha_i$  is the value of  $f_i$  (for  $0 \leq i \leq n$ ), and  $\beta_i$  (for  $1 \leq i \leq n+1$ ) is the argument of  $f_i$ .

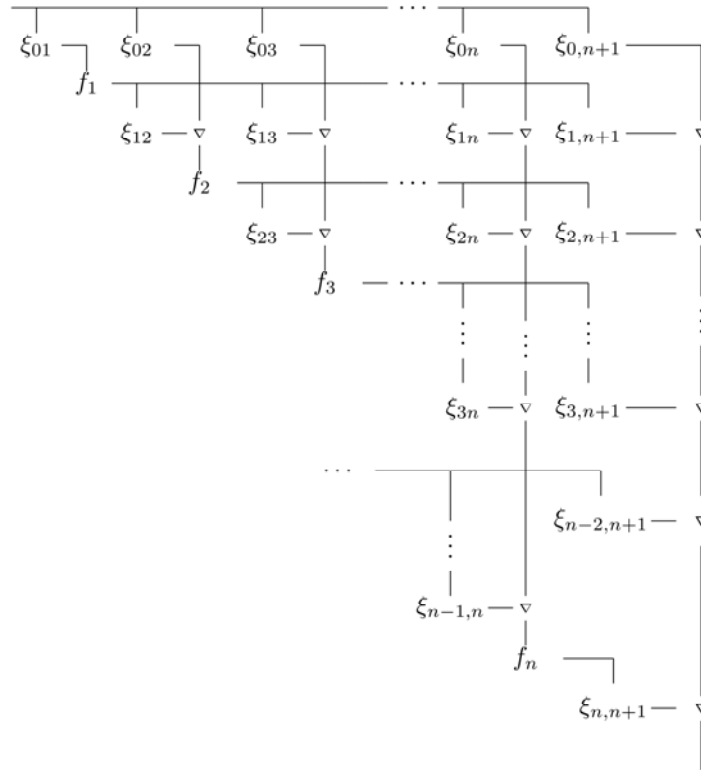


Fig. 2. Structure of an acyclic program, general case

**Relation of Acyclic NFs to Sequentially-Parallel NFs**

Considering features of acyclic NFs in general would require lots of technical details, so we will demonstrate these properties on some substantial examples.

**Statement 1.** Not every acyclic NF is sequential-parallel — see fig. 3 with parameters listed in table 1.

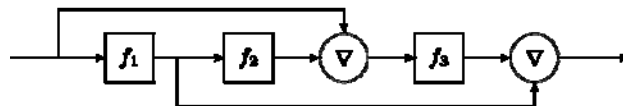


Fig 3. Example of an acyclic NF, that is not sequential-parallel

Table 1: Arities  $U_i, V_i$  and renamings  $\xi_{ij}$ .

$i \setminus j$	1	2	3	4	$V_i$
0	$id_{\{x,y\}}$	$\emptyset$	$id_{\{y\}}$	$\emptyset$	$\{x,y\}$
1	—	$id_{\{x\}}$	$\emptyset$	$id_{\{x\}}$	$\{x\}$
2	—	—	$id_{\{x\}}$	$\emptyset$	$\{x\}$
3	—	—	—	$id_{\{y\}}$	$\{y\}$
$U_i$	$\{x,y\}$	$\{x\}$	$\{x,y\}$	$\{x,y\}$	—

**Statement 2.** Every acyclic NF can be transformed into an equivalent sequentially-parallel NF. This transformation increases number of basic functions usages in the compositional term but does not introduce extra names. For the acyclic NF from the previous example, we have the following equivalent sequentially-parallel NF (see fig. 4):

$$f = f_1 \nabla \left( \left( \uparrow_{\{x\}}^{\{x,y\}} \nabla (f_1 \circ f_2) \right) \circ f_3 \right)$$

Apparently this sequentially-parallel NF has one extra usage of NF  $f_1$ .

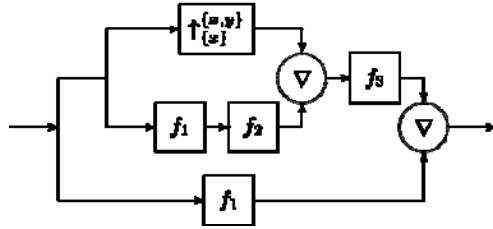


Fig 4. Sequentially-parallel NF equivalent to the acyclic NF

**Statement 3.** Every acyclic NF could be transformed, without introducing extra usages of basic functions to the compositional term but possibly introducing extra names, into an equivalent (up to extra names) chain NF.

Note that a weaker statement is trivial: indeed, every acyclic NF can be transformed to an equivalent sequential-parallel NF due to statement 2 (with extra usages of basic functions), and the later can be transformed to an equivalent chain NF due to the main theorem from [Parfirova 2010b]. To avoid extra usages of basic functions, however, we should transform acyclic NF to chain NF directly. For the given example, the equivalent chain NF (where parameters  $\xi_i$  and  $\zeta_i$  are shown in the table 2) is:

$$g = L_{\xi_1}^{\xi_1}(f_1) \circ L_{\zeta_2}^{\xi_2}(f_2) \circ L_{\zeta_3}^{\xi_3}(f_3) \circ \uparrow_{\{x,y\}}^{\{x,x',y\}}.$$

Table 2. Transformation to the chain NF

$i$	$\xi_i$	$\zeta_i$
1	$\{x \mapsto x, y \mapsto y\}$	$\{x \mapsto x, y \mapsto y\}$
2	$\{x' \mapsto x, y \mapsto y\}$	$\{x \mapsto x\}$
3	$\{x' \mapsto x, x \mapsto x', y \mapsto y\}$	$\{x \mapsto x, y \mapsto y\}$

This example is notable because it uses only one auxiliary name  $x'$ . On the other hand, we are interested in a general method of transformation acyclic NFs to chain NFs. Let us explain the main idea of the method. Suppose some  $f_i$  (including special case  $i = 0$  for the very input) has an output name, say  $x$ . Then the  $(i + 1)$ -th link of the chain NF should rename  $x$  to  $x_i$ . Except this, every (say,  $k$ -th) link must preserve all names of the form  $x_j$ ,  $0 \leq j < k$ , from its input. Therefore, the output of the  $k$ -th link consists of copies  $x_j$  of outputs of all previous subroutines and own outputs of  $f_k$ . Thus, if the  $j$ -th subroutine uses output  $x$  of the  $i$ -th subroutine when  $i < j - 1$  (i.e. if  $\xi_{ij}(y) = x$  for some name  $y$ ), the  $j$ -th link of the chain should rename  $x_i$  into  $y$  before applying  $f_k$ . The only remaining special issue is the last link that restores names from the form  $x_i$  and assembles the final result.

For the example above, this principle gives the following chain function with parameters shown in the table 3.

$$g = L_{\xi_1}^{\xi_1}(f_1) \circ L_{\zeta_2}^{\xi_2}(f_2) \circ L_{\zeta_3}^{\xi_3}(f_3) \circ L_{\zeta_4}^{\xi_4}(\text{id}_{\emptyset}) \circ \uparrow_{\{x,y\}}^{\{x_0,y_0,x_1,x_2,x,y\}}.$$

Table 3. Transformation to the chain NF

$i$	$\xi_i$	$\zeta_i$
1	$\{x_0 \mapsto x, y_0 \mapsto y\}$	$\{x \mapsto x, y \mapsto y\}$
2	$\{x_0 \mapsto x_0, y_0 \mapsto y_0, x_1 \mapsto x\}$	$\{x \mapsto x\}$
3	$\{x_0 \mapsto x_0, y_0 \mapsto y_0, x_1 \mapsto x_1, x_2 \mapsto x\}$	$\{x \mapsto x, y \mapsto y_0\}$
4	$\{x_0 \mapsto x_0, y_0 \mapsto y_0, x_1 \mapsto x_1, x_2 \mapsto x_2, x \mapsto x_1, y \mapsto y\}$	$\emptyset$



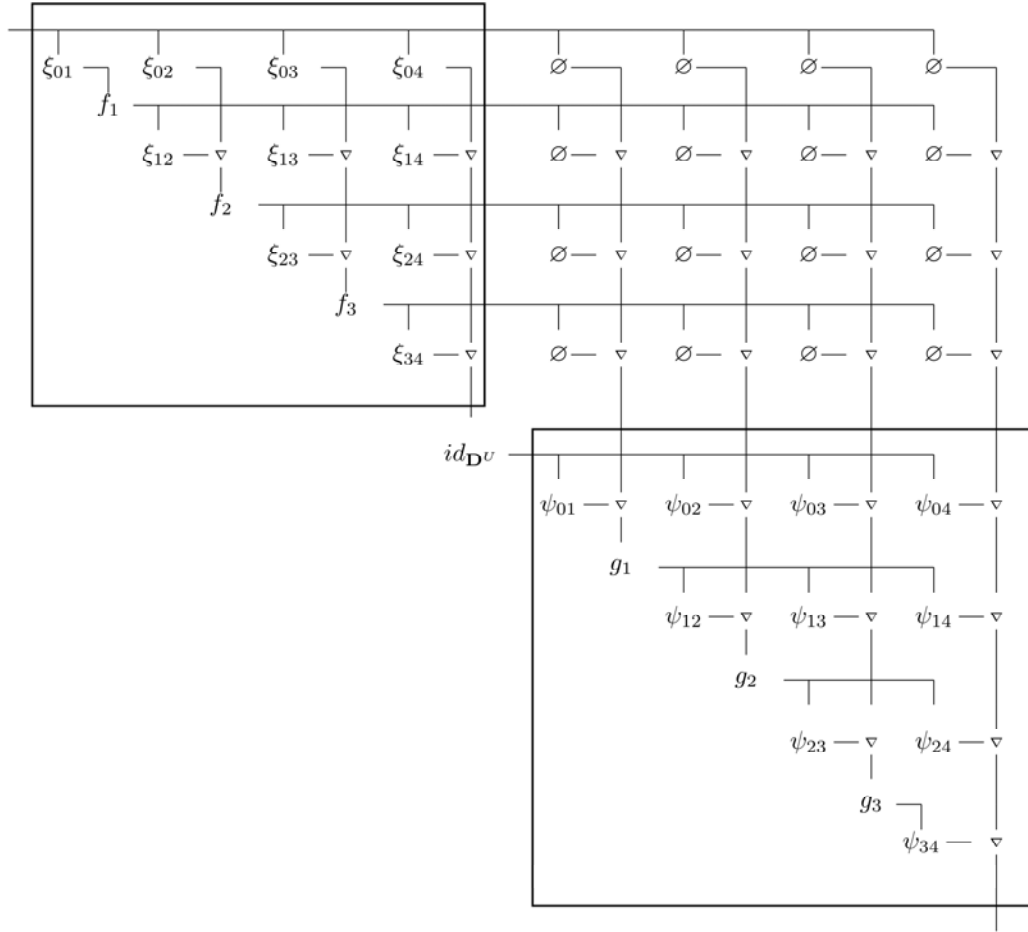


Fig. 5. Sequential composition of acyclic programs as an acyclic program

**Statement 4.** If  $f$  and  $g$  are both acyclic functions then their sequential composition  $f \circ g$  could be presented as an acyclic function as well. Let  $f = T_{V,U}^{\Xi}(f_1, \dots, f_m)$ ,  $g = T_{U,W}^{\Psi}(g_1, \dots, g_n)$ ; then (see fig. 5)  $f \circ g = h = T_{V,W}^Z(f_1, \dots, f_m, id_{D^U}, g_1, \dots, g_n)$  with the following matrix of renamings:

$$Z = \begin{bmatrix} \xi_{01} & \dots & \xi_{0,m+1} & \emptyset & \emptyset & \emptyset \\ & \ddots & \vdots & \emptyset & \emptyset & \emptyset \\ & & \xi_{m,m+1} & \emptyset & \emptyset & \emptyset \\ & & & \psi_{01} & \dots & \psi_{0,n+1} \\ & & & & \ddots & \vdots \\ & & & & & \psi_{n,n+1} \end{bmatrix} = \begin{bmatrix} \Xi & \emptyset \\ \emptyset & \Psi \end{bmatrix}.$$

In other words,

$$\zeta_{ij} = \begin{cases} \xi_{ij}, & \text{if } 0 \leq i < j \leq m+1, \\ \emptyset, & \text{if } 0 \leq i \leq m, m+2 \leq j \leq m+n+2, \\ \psi_{i-m-1, j-m-1}, & \text{if } m+1 \leq i < j \leq m+n+2. \end{cases}$$

From this example one can see that  $f \circ g$  can be transformed to an equivalent acyclic function  $h$  in such a way that the renaming matrix of  $h$  is very simple but one additional (though trivial) subroutine is introduced to  $h$ . It can be shown that there is another option: no extra subroutines are introduced to  $h$  but the renaming matrix is

not so simple. This should be regarded as an example of a common trade-off between simplicity of internal structure of a system vs. simplicity of interfaces between its subsystems.

---

## Conclusion

The composition  $T_{V,U}^{\Xi}$  described above models a sequential data and control flow between parts of a program that does not contain loops. Its particular cases are sequential programs, entirely parallel programs and various intermediate program structures that combine both parallel and sequential execution of statements. Finally, the model allows equivalent transformations by clear algebraic laws known from CP.

---

## Bibliography

- [Parfirova 2010a] T.S. Parfirova. Compositional model of sequential-parallel connections in information systems. Proceedings of the 10th international seminar "Discrete mathematics and its applications", Moscow State University, 198–201, 2010 (in Russian).
- [Parfirova 2010b] T.S. Parfirova. On reduction of sequential-parallel compositional programs to entirely sequential form. Bulletin of Kyiv National Taras Shevchenko University, series: Physical and mathematical sciences, 1, 132–137, 2010 (in Ukrainian).
- [Parfirova 2009] T.S. Parfirova. The notion of interaction from the perspective of the entity platform. Proceedings of the 6th international conference "Theoretical and applied aspects of program system development", Kyiv, 78–80, 2009 (in Russian).
- [Parfirova and Vinnik 2010] T. Parfirova, V. Vinnik. Copositional Model of Acyclic Programs. Proceedings of the International Conference on Computer Science and Engineering CSE'2010, Košice – Stará Ľubovňa, Slovakia (accepted for publication).
- [Redko 1978] V.N. Redko. Compositions of programs and compositional programming. Programming and Computer Software. 5, 3–24, 1978 (in Russian).
- [Redko 1998] V.N. Redko. Compositional structure of programmology. Cybernetics and Systems Analysis. 4, 47–66, 1998 (in Russian).
- [Redko 2008] V.N. Redko. Existential foundations of compositional paradigm. Cybernetics and Systems Analysis. 2, 3–12, 2008 (in Russian).
- [Redko et al. 2008] V.N. Redko, I.V. Redko, N.V. Gryshko. Programmological foundations of entity platform. Problems of Programming. 2–3, 75–83, 2008 (in Russian).
- [Nikitchenko 1999] N.S. Nikitchenko. Compositional-nominative approach to clarifying the notion of program. Problems of Programming. 1, 16–31, 1999 (in Russian).
- [Nikitchenko 2009] N.S. Nikitchenko. Compositional-nominative aspects of address programming. Cybernetics and Systems Analysis. 6, 24–35, 2009 (in Russian).

---

## Authors' Information

**Tetiana Parfirova** – Kiev National Taras Shevchenko University, faculty of Cybernetics, 01601, Volodymyrska str, 64, Kiev, Ukraine; e-mail [tetiana.parfirova@gmail.com](mailto:tetiana.parfirova@gmail.com)

Major Fields of Scientific Research: Theory of programming, formal semantics of programming languages, educational environments, e-learning systems.

**Vadim Vinnik** – Kiev National Taras Shevchenko University, faculty of Cybernetics, 01601, Volodymyrska str, 64, Kiev, Ukraine; e-mail [vadim.vinnik@gmail.com](mailto:vadim.vinnik@gmail.com)

Major Fields of Scientific Research: Theory of programming, formal semantics of programming languages, program specification and verification, complex systems modelling and simulation





## ITHEA INTERNATIONAL SCIENTIFIC SOCIETY

The ITHEA International Scientific Society (**ITHEA ISS**) a successor of the international scientific co-operation organized within 1986-1992 by international workgroups (**IWG**) researching the problems of data bases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the international scientific group of Data Base Intellectualization (**IWGDBI**) integrating the possibilities of databases with the creative process support tools. The leaders of the IWGDBI were Prof. Victor Gladun (Ukraine) and Prof. Romyana Kirkova (Bulgaria). Starting from 1992 till now the international scientific co-operation has been organized by the Association of Developers and Users of Intellectualized Systems (**ADUIS**), Ukraine. It has played a significant role for uniting the scientific community working in the area of the artificial intelligence.

To extend the possibilities for international scientific collaboration in all directions of informatics by wide range of concrete activities, in 2002 year, the Institute for Information Theories and Applications FOI ITHEA (**ITHEA**) has been established as an international nongovernmental organization.

**ITHEA Scientific Council** includes:

*Honorable Chair Persons:* Victor Gladun (Ukraine) and Romyana Kirkova (Bulgaria);

*Chairman:* Levon Aslanyan (Armenia);

*Members:* Adil Timofeev (Russia), Alexander Kleshchev (Russia), Alexander Kuzemin (Ukraine), Alexander Palagin (Ukraine), Alexey Voloshyn (Ukraine), Arkadij Zakrevskij (Belarus), Constantin Gaidric (Moldova), Hasmik Sahakyan (Armenia), Ilia Mitov (Bulgaria), Juan Castellanos (Spain), Koen Vanhoof (Belgie), Krassimir Markov (Bulgaria), Larisa Zainutdinova (Russia), Laura Ciocoiu (Romania), Luis Fernando de Mingo (Spain), Martin Mintchev (Canada), Nikolay Zagoruiko (Russia), Peter Stanchev (USA), Stefan Karastanev (Bulgaria), Tatyana Gavrilova (Russia), Vladimir Lovitskii (GB), Vladimir Ryazanov (Russia).

**ITHEA Scientific Sections:** Business Informatics (Koen Vanhoof), Cooperative Scientific Projects (Hasmik Sahakyan), Decision Support (Alexey Voloshyn), Information Interaction (Stefan Karastanev), Intelligent Engineering Systems (Martin Mintchev), Logical Inference and Design (Arkadij Zakrevskij), Modern (e-) Learning (Larisa Zaynutdinova), Natural Computing (Juan Castellanos), Pattern Recognition (Vladimir Ryazanov), Philosophy and Methodology of Informatics (Krassimir Markov), Scientific Innovation Management (Luis Fernando de Mingo), Software Engineering (Ilia Mitov), Theoretical Informatics (Levon Aslanyan).

ITHEA is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc. The achieved results are remarkable. The ITHEA became world-wide known scientific organization. One of the main activities of the ITHEA is building an International Scientific Society aimed to unite researches from all over the world who are working in the area of informatics.

Now, the **ITHEA International Scientific Society** is joined by more than **2200** members from **44** countries all over the world: Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia and Montenegro, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.



## ASSOCIATION OF DEVELOPERS AND USERS OF INTELLIGENT SYSTEMS

**ADUIS** consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of **ADUIS** is to contribute to the development and application of the artificial intelligence methods and techniques. The efforts of scientists engaged in **ADUIS** are concentrated on the following problems: expert system design; knowledge engineering; knowledge discovery; planning and decision making systems; cognitive models designing; human-computer interaction; natural language processing; methodological and philosophical foundations of AI.

Association has long-term experience in collaboration with teams, working in different fields of **research and development**. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties. Some thousands of high precise prognoses have been done in collaboration with chemists and material scientists of Russia and USA.

Association can help **businessmen** to find out conditions for successful investment taking into account region or field peculiarities as well as to reveal user's requirements on technical characteristics of products being sold or manufactured.

**Physicians** can be equipped with systems, which help in diagnosing or choosing treatment methods, in forming multi-parametric models that characterize health state of population in different regions or social groups.

**Sociologists, politicians, managers** can obtain the Association's help in creating generalized multi-parametric "portraits" of social groups, regions, enterprise groups. Such "portraits" can be used for prognostication of voting results, progress trends, and different consequences of decision making as well.

Association provides a useful guide in technical diagnostics, ecology, geology, and genetics.

ADUIS has at hand a broad range of high-efficiency original methods and program tools for solving analytical problems, such as knowledge discovery, classification, diagnostics, and prognostication.

ADUIS unites the creative potential of highly skilled scientists and engineers

Since 1992 **ADUIS** holds regular conferences and workshops with wide participation of specialists in AI and users of intelligent systems. The proceedings of the conferences and workshops are published in scientific journals.

**ADUIS** cooperates through its foreign members with organizations that work on AI problems in Russia, Belarus, Moldova, Georgia, Bulgaria, Czech Republic, Germany, Great Britain, Hungary, Poland, etc. **ADUIS** is the collective member of the European Coordinating Committee for Artificial Intelligence (ECCAI).

[www.aduis.com.ua](http://www.aduis.com.ua)

For contacts: V.M.Glushkov Institute of Cybernetics; National Academy of Science of Ukraine;

Prospect Akademika Glushkova, 40, 03680 GSP Kiev-187, Ukraine;

Phone: (380+44) 5262260; Fax: (380+44) 5263348; E-mail: [glad@aduis.kiev.ua](mailto:glad@aduis.kiev.ua)

