

# БИЗНЕС- ИНФОРМАТИКА

МЕЖДИСЦИПЛИНАРНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

## СОДЕРЖАНИЕ

### *Анализ данных и интеллектуальные системы*

*М.И. Коломейченко, А.М. Чеповский*

Визуализация и анализ графов больших размеров .....7

*М.О. Ланин*

Автоматический поиск опорных элементов  
на документах полужесткой структуры ..... 17

*М.А. Орлов*

Алгоритм формирования  
многокритериальной стратификации .....24

### *Информационные системы и технологии в бизнесе*

*И.Н. Фомин, Н.В. Сердюкова*

Расчетная модель измерения электроэнергии  
в информационных биллинговых системах.....36

*В.Г. Чеботарёв, А.И. Громов*

Автоматизация процесса обучения .....45

### *Математические методы и алгоритмы решения задач бизнес-информатики*

*Р. Malyzhenkov, T. Babkina, A. Sergeev*

Flexible organizational forms design  
based on a transaction approach.....53

### *Программная инженерия*

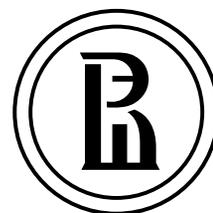
*D. Pashchenko, A. Blinov*

Standardization in software production  
at the corporate level: Results of research in CIS .....63

### *Математическое моделирование социально-экономических процессов*

*T. Bogdanova, T. Yakovets*

World demographic situation from the perspective  
of global demographic balance .....72



*Издатель:*

Национальный  
исследовательский университет  
«Высшая школа экономики»

**Подписной индекс  
в каталоге агентства  
«Роспечать» –72315**

Выпускается ежеквартально

*Журнал включен в Перечень  
российских рецензируемых  
научных журналов,  
в которых должны быть  
опубликованы основные научные  
результаты диссертаций  
на соискание ученых степеней  
доктора и кандидата наук*

*Главный редактор  
А.О. Голосов*

*Заместители главного редактора  
А.Р. Горбунов, Д.В. Исаев*

*Научный редактор  
Н.Н. Лычкина*

*Технический редактор  
В.И. Осипов*

*Дизайн обложки  
С.Н. Борисова*

*Компьютерная верстка  
О.А. Богданович*

*Администратор веб-сайта  
Д.С. Проценко*

Журнал рекомендован ВАК  
для научных публикаций

*Адрес редакции:*

105187, г. Москва, ул. Кирпичная, д. 33

Тел./факс: +7 (495) 771-32-38

<http://bijournal.hse.ru>

E-mail: [bijournal@hse.ru](mailto:bijournal@hse.ru)

За точность приведенных сведений  
и содержание данных,  
не подлежащих открытой публикации,  
несут ответственность авторы

**При перепечатке ссылка на журнал  
«Бизнес-информатика» обязательна**

*Тираж 500 экз.*

Отпечатано в типографии НИУ ВШЭ  
г. Москва, Кочновский проезд, 3

© Национальный  
исследовательский университет  
«Высшая школа экономики»

# О ЖУРНАЛЕ

«**Б**изнес-информатика» – рецензируемый междисциплинарный научный журнал, выпускаемый с 2007 года Национальным исследовательским университетом «Высшая школа экономики» (НИУ ВШЭ). Администрирование журнала осуществляется факультетом бизнес-информатики НИУ ВШЭ.

Миссия журнала – развитие бизнес-информатики как новой области информационных технологий и менеджмента. Журнал осуществляет распространение последних разработок технологического и методологического характера, способствует развитию соответствующих компетенций, а также обеспечивает возможности для дискуссий в области применения современных информационно-технологических решений в бизнесе, менеджменте и экономике.

Журнал публикует статьи по следующей тематике:

- ◆ корпоративные информационные системы;
- ◆ информационные технологии в бизнесе;
- ◆ организационные и управленческие проблемы создания и внедрения информационных систем;
- ◆ математическое моделирование социально-экономических процессов;
- ◆ методы анализа информации;
- ◆ интеллектуальные системы и управление знаниями в бизнесе;
- ◆ информационные сети и телекоммуникации;
- ◆ программная инженерия;
- ◆ информационная безопасность;
- ◆ электронный бизнес;
- ◆ инновации и бизнес в сфере информационных технологий;
- ◆ стандартизация, сертификация и качество;
- ◆ правовые вопросы бизнес-информатики;
- ◆ обучение в сфере бизнес-информатики.

В соответствии с решением президиума Высшей аттестационной комиссии Российской Федерации с 2010 года журнал включен в Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук.

Журнал выпускается ежеквартально и распространяется как в печатном виде, так и в электронной форме.

Журнал «Бизнес-информатика» зарегистрирован в «Роскомнадзоре».  
Свидетельство ПИ № ФС 7752404 от 28 декабря 2012 г.

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

### ГЛАВНЫЙ РЕДАКТОР

**ГОЛОСОВ Алексей Олегович** —  
кандидат технических наук, Президент компании  
«ФОРС — Центр разработки»

### ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА

**ГОРБУНОВ Алексей Рэмович** —  
кандидат экономических наук, старший научный сотрудник,  
Института США и Канады РАН

**ИСАЕВ Дмитрий Валентинович** —  
кандидат экономических наук, доцент кафедры бизнес-  
анализа, факультет бизнес-информатики, Национальный  
исследовательский университет «Высшая школа экономики»

### ЧЛЕНЫ РЕДКОЛЛЕГИИ

**АБДУЛЬРАБ Абиб** —  
PhD, профессор департамента математики и программной  
инженерии, Национальный институт прикладных наук, Руан,  
Франция

**АВДОШИН Сергей Михайлович** —  
кандидат технических наук, заведующий кафедрой управления  
разработкой программного обеспечения, заведующий  
отделением программной инженерии, Национальный  
исследовательский университет «Высшая школа экономики»

**АЛЕСКЕРОВ Фуад Тагиевич** —  
доктор технических наук, профессор, руководитель  
департамента математики, факультет экономики, Национальный  
исследовательский университет «Высшая школа экономики»

**БАБКИН Эдуард Александрович** —  
кандидат технических наук, PhD, заведующий кафедрой  
информационных систем и технологий, факультет бизнес-  
информатики и прикладной математики (Нижегород),  
Национальный исследовательский университет  
«Высшая школа экономики»

**БАЙЕР Алекс** —  
PhD, Директор KAFAN FX Information Services, Нью-Йорк, США

**БАРАНОВ Александр Павлович** —  
доктор физико-математических наук, профессор,  
заведующий кафедрой информационной безопасности, факуль-  
тет бизнес-информатики, Национальный исследовательский  
университет «Высшая школа экономики»

**БЕККЕР Йорг** —  
PhD, проректор, профессор, директор Европейского  
исследовательского центра в области информационных систем  
(ERCIS) Мюнстерского университета, Мюнстер, Германия

**БЕЛОВ Владимир Викторович** —  
доктор технических наук, профессор кафедры вычислительной  
и прикладной математики, факультет вычислительной техники,  
Рязанский государственный радиотехнический университет.

**ГРИБОВ Андрей Юрьевич** —  
кандидат экономических наук, Генеральный директор  
компании «КиберПлат»

**ГРОМОВ Александр Игоревич** —  
кандидат химических наук, профессор, заведующий кафедрой  
моделирования и оптимизации бизнес-процессов, факультет  
бизнес-информатики, Национальный исследовательский  
университет «Высшая школа экономики»

**ГУРВИЧ Владимир Александрович** —  
PhD, приглашенный профессор и исследователь,  
Центр исследования операций, Ратгерский университет  
(Университет Нью-Джерси), США

**ДЖЕЙКОБС Лоренц** —  
PhD, профессор медицинского факультета, Университет  
Цюриха, Швейцария

**ЗАНДКУЛЬ Курт** —  
PhD, заведующий кафедрой информационных систем  
для бизнеса, институт информатики, факультет информатики  
и электротехники, Университет Ростока, Германия

**ИЛЬИН Николай Иванович** —  
доктор технических наук, профессор, заместитель начальника  
Управления специальной связи, Федеральная служба охраны  
Российской Федерации (ФСО России)

**КАЛЯГИН Валерий Александрович** —  
доктор физико-математических наук, профессор,  
заведующий кафедрой прикладной математики и информатики,  
факультет бизнес-информатики и прикладной математики  
(Нижегород), Национальный исследовательский  
университет «Высшая школа экономики»

**КАМЕННОВА Мария Сергеевна** —  
кандидат технических наук, директор компании «ИДС Шер»,  
Россия и страны СНГ

**КУЗНЕЦОВ Сергей Олегович** —  
доктор физико-математических наук, профессор, заведующий  
кафедрой анализа данных и искусственного интеллекта,  
заведующий отделением прикладной математики, Национальный  
исследовательский университет «Высшая школа экономики»

**МАЛЬЦЕВА Светлана Валентиновна** —  
доктор технических наук, профессор, заведующий кафедрой  
инноваций и бизнеса в сфере информационных технологий,  
декан факультета бизнес-информатики, Национальный  
исследовательский университет «Высшая школа экономики»

**МЕЙОР Питер** —  
PhD, заместитель директора консультативной группы  
по радиокommunikациям, Международный телекоммуникаци-  
онный союз (ITU), заместитель руководителя Комиссии ООН  
по науке и технологиям, Женева, Швейцария

**МИРКИН Борис Григорьевич** —  
доктор технических наук, профессор кафедры анализа данных  
и искусственного интеллекта, отделение прикладной мате-  
матики и информатики, Национальный исследовательский  
университет «Высшая школа экономики»

**МОТТЛЬ Вадим Вячеславович** —  
доктор технических наук, профессор кафедры  
информационной безопасности, факультет кибернетики,  
Тульский государственный университет

**ПАЛЬЧУНОВ Дмитрий Евгеньевич** —  
доктор физико-математических наук, заведующий кафедрой  
общей информатики, факультет информационных технологий,  
Новосибирский государственный университет

**ПАРДАЛОС Панайот (Панос)** —  
PhD, почетный профессор, директор центра прикладной  
оптимизации, департамент промышленной и системной  
инженерии, Университет Флориды, США

**СИЛАНТЬЕВ Альберт Юрьевич** —  
доктор технических наук, профессор кафедры  
информационных бизнес систем, Институт информационных  
бизнес-систем, Национальный исследовательский  
технологический университет «МИСиС»

**ТАРАТУХИН Виктор Владимирович** —  
кандидат технических наук, PhD, руководитель научной группы  
Европейского исследовательского центра в области  
информационных систем (ERCIS) Мюнстерского  
университета, Мюнстер, Германия

**УЛЬЯНОВ Михаил Васильевич** —  
доктор технических наук, профессор кафедры управления  
разработкой программного обеспечения, отделение  
программной инженерии, Национальный исследовательский  
университет «Высшая школа экономики»

**ШАЛКОВСКИЙ Алексей Геннадьевич** —  
кандидат технических наук, проректор, директор Института  
информационных технологий, Национальный  
исследовательский университет «Высшая школа экономики»

## CONTENTS

### *Data analysis and intellectual systems*

*M. Kolomeychenko, A. Chepovskiy*

Huge graph visualization and analysis .....7

*Mikhail Lanin*

Automatic detection of reference elements  
on semi-structured document images ..... 17

*Mikhail Orlov*

An algorithm for multicriteria stratification ..... 24

### *Information systems and technologies in business*

*I. Fomin, N. Serdyukova*

Computational model for measuring electric power  
in meter-to-cash systems ..... 36

*V. Chebotarev, A. Gromov*

Automation of education process ..... 45

### *Mathematical methods and algorithms for business informatics problems*

*P. Malyzhenkov, T. Babkina, A. Sergeev*

Flexible organizational forms design  
based on a transaction approach..... 53

### *Software engineering*

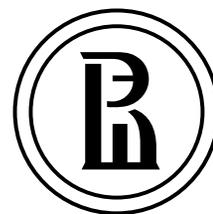
*D. Pashchenko, A. Blinov*

Standardization in software production  
at the corporate level: Results of research in CIS ..... 63

### *Mathematical modeling of economic and social processes*

*T. Bogdanova, T. Yakovets*

World demographic situation from the perspective  
of global demographic balance ..... 72



*Publisher:*

National Research University –  
Higher School of Economics

**Subscription index  
in the «Rospechat» catalogue –  
72315**

The journal is published quarterly

*The journal is included  
into the list of peer reviewed  
scientific editions established  
by the Supreme Certification  
Commission of the Ministry  
of Education and Science  
of the Russian Federation*

*Editor-in-Chief:*

**A. Golosov**

*Deputies Editor-in-Chief:*

**A. Gorbunov, D. Isaev**

*Scientific Editor:*

**N. Lychkina**

*Technical Editor:*

**V. Osipov**

*Design:*

**S. Borisova**

*Computer Making-up:*

**O. Bogdanovich**

*Website Administration:*

**D. Protsenko**

*Address:*

33, Kirpichnaya str., Moscow,  
105187, Russian Federation

Tel./fax: +7 (495) 771-32-38

<http://bijournal.hse.ru>

E-mail: [bijournal@hse.ru](mailto:bijournal@hse.ru)

*Circulation – 500 copies*

Printed in HSE Printing House  
3, Kochnovsky proezd, Moscow,  
Russian Federation

© National Research University –  
Higher School of Economics

# ABOUT THE JOURNAL

**B**usiness Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by National Research University – Higher School of Economics (HSE), Moscow, Russian Federation. The journal is administered by Faculty of Business Informatics.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the areas of, but not limited to:

- ◆ Corporate information systems
- ◆ Information technologies in business
- ◆ Organizational and managerial problems of information systems development and implementation
- ◆ Mathematical modeling of economic and social processes
- ◆ Methods of information analysis
- ◆ Intellectual systems and knowledge management in business
- ◆ Information networks and telecommunications
- ◆ Software engineering
- ◆ Information security
- ◆ Electronic business
- ◆ Innovations and business in the sphere of information technologies
- ◆ Standardization, certification and quality
- ◆ Legislation in the field of business informatics
- ◆ Education in the field of business informatics.

Since 2010 the journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Ministry of Education and Science of the Russian Federation.

The journal is published quarterly and distributed both in printed and electronic forms.

---

---

## EDITORIAL BOARD

### *EDITOR-IN-CHIEF*

**Dr. Alexey GOLOSOV** –

President of FORS Development Center, Russian Federation

### *DEPUTY EDITOR-IN-CHIEF*

**Dr. Alexey GORBUNOV** –

Senior Researcher, Institute of US and Canada Studies

**Dr. Dmitry ISAEV** –

Associate Professor, Department of Business Analytics, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

### *EDITORIAL BOARD*

**Prof. Habib ABDULRAB** –

Professor, Mathematical and Software Engineering Department, National Institute of Applied Sciences – Institut national des sciences appliquées de Rouen (INSA de Rouen), Rouen, France

**Dr. Sergey AVDOSHIN** –

Head of Software Management Department, Head of School of Software Engineering, National Research University Higher School of Economics, Russian Federation

**Prof. Fuad ALESKEROV** –

Head of Department of Mathematics, Faculty of Economics, National Research University Higher School of Economics, Russian Federation

**Dr. Eduard BABKIN** –

Head of Department of Information Systems and Technologies, Faculty of Business Informatics and Applied Mathematics (Nizhny Novgorod), National Research University Higher School of Economics, Russian Federation

**Dr. Alex BAYER** –

Head of KAFAN FX Information Services, New York, USA

**Prof. Alexander BARANOV** –

Head of Department of Information Security Management, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

**Prof. Jorg BECKER** –

Vice Rector, Professor, Director of European Research Center for Information Systems (ERCIS), University of Munster, Germany

**Prof. Vladimir BELOV** –

Professor, Department of Computational and Applied Mathematics, Faculty of Computer Engineering, Ryazan State Radio Engineering University, Russian Federation

**Dr. Andrey GRIBOV** –

Director General, CyberPlat Company, Russian Federation

**Prof. Alexander GROMOV** –

Head of Department of Modeling and Business Process Optimization, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

**Prof. Vladimir GURVICH** –

Invited Professor and Researcher, Rutgers Center for Operations Research, Rutgers, The State University of New Jersey, USA

**Prof. Laurence JACOBS** –

Professor, Medical School, University of Zurich, Switzerland

**Prof. Kurt SANDKUHL** –

Head of Department of Business Information Systems, Institute of Computer Science, Faculty of Computer Science and Electrical Engineering, University of Rostock, Germany

**Dr. Nikolay ILYIN** –

Deputy Head, Administration of Special Communication, Federal Security Guard, Russian Federation

**Prof. Valery KALYAGIN** –

Head of Department of Applied Mathematics and Informatics, Faculty of Business Informatics and Applied Mathematics (Nizhny Novgorod), National Research University Higher School of Economics, Russian Federation

**Dr. Maria KAMENNOVA** –

director, IDS Scheer Russia and CIS Countries, Russian Federation

**Prof. Sergey KUZNETSOV** –

Head of Department of Data Analysis and Artificial Intelligence, Head of School of Applied Mathematics and Information Science, National Research University Higher School of Economics, Russian Federation

**Prof. Svetlana MALTSEVA** –

Head of Department of Innovation and Business in Information Technologies, Acting Dean of Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

**Dr. Peter MAJOR** –

Vice-chairman, Radiocommunication Advisory Group of International Telecommunication Union (ITU), vice-chairman of the UN Commission on Science and Technology for Development (CSTD), Geneva, Switzerland

**Prof. Boris MIRKIN** –

Professor of Department of Data Analysis and Artificial Intelligence, School of Applied Mathematics and Information Science, National Research University Higher School of Economics, Russian Federation

**Prof. Vadim MOTIL** –

Professor, Department of Information Security Management, Faculty of Cybernetics, Tula State University, Russian Federation

**Prof. Dmitry PALCHUNOV** –

Head of Department of General Informatics, Faculty of Information Technologies, Novosibirsk State University, Russian Federation

**Prof. Panagote (Panos) PARDALOS** –

Distinguished Professor and University of Florida Research Foundation Professor, Director of Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, USA

**Dr. Albert SILANTYEV** –

Professor, Department of Information Business Systems, Institute of Information Business Systems, National University of Science and Technology «MISIS», Russian Federation

**Dr. Victor TARATOUKHIN** –

Managing Director European Research Center for Information Systems (ERCIS) Competence Center ERP, Head of ERCIS Lab. Russia, University of Munster, Germany

**Prof. Mikhail ULYANOV** –

Professor of Software Management Department, School of Software Engineering, National Research University Higher School of Economics, Russian Federation

**Dr. Alexey SHALCOVSKY** –

Vice Rector, Director of Institute of Information Technologies, National Research University Higher School of Economics, Russian Federation

# ВИЗУАЛИЗАЦИЯ И АНАЛИЗ ГРАФОВ БОЛЬШИХ РАЗМЕРОВ

## **М.И. КОЛОМЕЙЧЕНКО**

выпускник магистратуры,

Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: maxim.kolomeychenko@mail.ru

## **А.М. ЧЕПОВСКИЙ**

кандидат технических наук, доцент кафедры информационной безопасности, факультет бизнес-информатики, Национальный исследовательский университет «Высшая школа экономики»; профессор кафедры прикладной математики и моделирования систем, Институт коммуникаций и медиабизнеса, Московский государственный университет печати им. Ивана Федорова

Адрес: 127550, г. Москва, ул. Прянишникова, д. 2а

E-mail: achepovskiy@hse.ru

Задача визуализации графов больших размеров возникает в различных областях социологии и маркетинга. Актуальность данной работы определяется потребностью в программном комплексе для анализа и визуализации таких графов. В работе приводится анализ нескольких программных продуктов и выделяются их недостатки: отсутствие кроссплатформенности и специализированных графовых хранилищ, а также невозможность работы с графами больших размеров.

Приводится детальное описание общей архитектуры разработанного программного обеспечения и каждого модуля в отдельности, способы взаимодействия основных модулей. Для хранения графов используется разработанное специализированное графовое хранилище, позволяющее обрабатывать графы, имеющие порядка 100 миллионов вершин и нескольких миллиардов связей. Также представлено описание основных принципов организации хранилища. Использование собственной файловой системы обеспечивает отсутствие дополнительных системных вызовов при работе с хранилищем и отсутствие сложной системы адресации и лишних механизмов, что приводит к избавлению от дополнительных накладных расходов, связанных с организацией хранения данных.

Кроме того, присутствует описание методики работы модуля визуализации данных, используемых структур данных и алгоритмов машинной графики, которые позволяют работать с графами, состоящими из нескольких миллионов вершин, в режиме реального времени.

Отдельно стоит отметить широкий набор алгоритмов автоматического размещения графов: случайное размещение, круговое размещение, круговое покомпонентное размещение, размещение «навлиний хвост», размещения «одна и две линии темы», размещения, основанные на алгоритме выделения сообществ или на алгоритме оценки связности. Приводится детальное описание каждого приведенного выше алгоритма.

Особое внимание стоит уделить используемым методам анализа графа. Разработаны алгоритмы выделения сообществ в социальных сетях, оценки связности графа, поиск кратчайших путей между любой парой вершин, объединение и пересечение графов и многое другое.

Ключевой особенностью всех приведенных в работе алгоритмов является возможность работы с графами больших размеров.

**Ключевые слова:** граф, анализ графа, визуализация графа, графовое хранилище, авторазмещение.

**Введение**

**П**ри анализе графов социальных сетей [1] возникает задача визуального представления их структур. Актуальность работы определяется потребностью в программном комплексе для анализа и визуализации графов больших размеров, возникающей в задачах социологического и маркетингового анализа.

Представленное программное обеспечение предоставляет широкие возможности для визуализации и анализа сетей больших размеров. Область применения обширна [5] и затрагивает многие смежные дисциплины, такие как социология, психология, политология, маркетинг и т.д. Существуют такие программные продукты для анализа графов, как i2 Analyst’s Notebook [9], Sentinel Visualizer [12], CrimeLink [8], Xanalis Link Explorer [14] и Tom Sawyer Software [13]. Несмотря на некоторые различия в деталях, по предоставляемой функциональности и назначению эти системы визуализации во многом схожи.

i2 Analyst’s Notebook, CrimeLink, Sentinel Visualizer и Xanalis Link Explorer являются программными продуктами, предназначенными для анализа систем взаимосвязанных объектов и изучения динамики последовательных событий. Tom Sawyer Software представляет собой набор библиотек для создания инструментов визуализации и анализа сетей из различных предметных областей.

В перечисленных выше системах наблюдается ряд недостатков: отсутствие платформенно-независимых решений, отсутствие собственных специализированных хранилищ, а также то, что системы не ориентированы на работу с графами больших размеров.

**1. Архитектура программного комплекса анализа графов**

Программный комплекс предназначен для визуального анализа графов взаимодействующих объектов (например, социальной сети) и автоматизированного размещения вершин и связей при визуализации в соответствии с задаваемыми образительными соглашениями.

Основные компоненты программного комплекса и взаимодействие модулей, составляющих программный комплекс, представлены на *рис. 1*.

Импорт данных в приложение осуществляется с помощью модуля конвертации данных. Реализо-

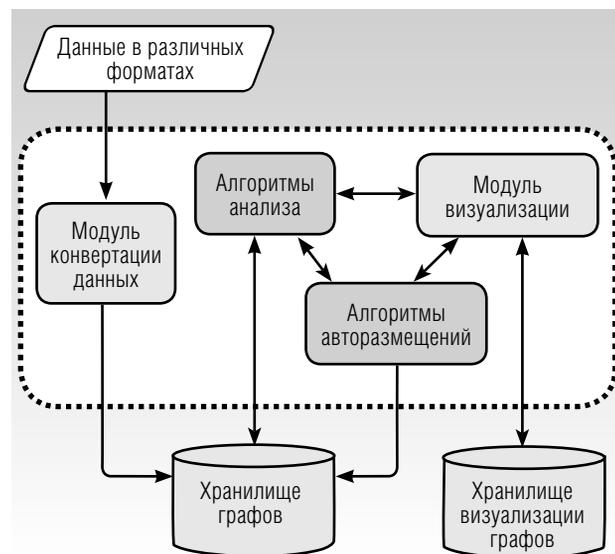


Рис 1. Архитектура программного комплекса

вана обработка данных из таблиц Excel, файлов с расширениями: doc, html, xml, реляционной БД, или текстовых файлов специального формата, содержащих структурированные данные. Конвертер выделяет из файлов табличные данные и преобразовывает их в граф, который сохраняется в файловой системе хранилища графа.

Принципиальной основой архитектуры программного комплекса является наличие двух режимов хранения графа:

- ◆ базовое хранение, предназначенное для хранения графов большого размера (миллионы вершин и ребер);
- ◆ локальное хранение, предназначенное для хранения графов, которые являются подграфами, полученными запросами к базовым графам и результатами обработки клиентским приложением.

В качестве базовой структуры хранилища используется специализированная файловая система [4]. Данная система реализуется внутри некоторого файла стандартной файловой системы компьютера, поэтому под логическим файлом будем понимать файл, принадлежащий данной специализированной системе и состоящий из записей переменной длины.

Использование собственной файловой системы обеспечивает отсутствие дополнительных системных вызовов при работе с файловой системой. Отсутствие сложной системы адресации (структуры каталогов) и лишних механизмов (права доступа) приводит к отсутствию дополнительных накладных расходов на организацию хранения данных.

Собственная файловая система позволяет организовывать управление буферизацией и хранением данных, которое не зависит от особенностей операционной системы.

Для более быстрого выполнения операций поиска вершин по конкретным значениям атрибутов в хранилище присутствует специализированный индекс по атрибутам вершин. Данный индекс также использует собственную копию специализированной файловой системы.

Используемые структуры данных позволяют проводить эффективные объединение и пересечение графов. Данная структура хранилища и предусмотренные операции были протестированы на графах, имеющих порядка 100 миллионов вершин и нескольких миллиардов связей.

К каждому конкретному сохраненному графу может быть привязано любое количество вариантов визуализации локального графа, которые сохраняются в хранилищах вариантов визуализации. Под вариантом визуализации понимается расположение вершин и связей между собой, иконки всех вершин, способы визуализации всех вершин, иконки всех атрибутов и их стили, стили всех связей и координаты их точек изломов.

Схема визуализации для графа сохраняется в бинарный файл. Сначала записывается строковый уникальный идентификатор хранилища для того, чтобы ввести однозначное соответствие схемы визуализации и графа. Затем в файл записывается вся коллекция иконок вершин в формате <имя иконки> <изображение иконки>. Далее в файл записывается вся коллекция типов атрибутов в формате <имя> <цвет> <флаг отображения> <имя иконки> <иконка> <порядковый номер при визуализации> <флаг отображения иконки> <флаг отображения имени> <тип выравнивания текстового значения> <тип хранимого значения>. Затем в файл записывается вся коллекция типов связей в формате <имя типа связи> <цвет типа связи> <ширина линии> <стиль линии> <направление линии>. Далее записываются группировки вершин и связей: для каждой групповой вершины записывается ее идентификатор и массив идентификаторов вершин, из которых она состоит; для каждой групповой связи записывается ее идентификатор и массив идентификаторов связей, из которых она состоит. Запись групповых вершин и связей происходит с учетом последовательности и вложенности группировок. Затем последовательно сохраняется визуализация всех вершин в формате

<идентификатор вершины> <координата x> <координата y> <идентификатор способа визуального представления вершины> <дополнительные параметры конкретного визуального представления>. После сохраняется визуализация всех связей в формате <идентификатор связи> <идентификатор вершины начала> <идентификатор вершины конца> <стиль> <ширина> <отступ первого сегмента излома> <направление>. Далее идет запись всех изломов связей в формате <идентификатор вершины начала связей> <идентификатор вершины конца связей> <величина отступа между параллельно идущими связями> <количество изломов> <массив координат всех изломов>.

Ниже отдельно приведено описание модуля визуализации и модулей, реализующих алгоритмы автотематического размещения и анализа графа.

## 2. Модуль визуализации графа

Спроектированы эффективные структуры данных, которые позволили создать развитую систему редактирования и анализа без ущерба производительности и с допустимым объемом занимаемой памяти. Для быстрого доступа к вершинам и связям используются сбалансированные бинарные деревья поиска. Для быстрого доступа к инцидентным связям вершины используется еще одна структура – сбалансированное бинарное дерево поиска, ключами в котором являются указатели на вершины, а хранимым значением для каждого ключа (вершины) – сбалансированное бинарное дерево поиска, ключами в котором являются указатели на связи. В данной структуре для каждой вершины хранится набор инцидентных связей. В табл. 1 приведен анализ эффективности базовых операций с выбранной структурой при работе с графом, полностью загруженным в оперативную память. Пусть  $V$  – множество вершин графа,  $E$  – множество ребер графа.

Таблица 1.  
Асимптотика операций при работе с графом

Операция	Асимптотика операций
Добавление связи	$O(\log( E ))$
Удаление связи	$O(\log( E ))$
Добавление вершины	$O(\log( V ))$
Удаление вершины	$O(\log( V ))$
Поиск вершины	$O(\log( V ))$
Поиск связи	$O(\log( E ))$

Объем занимаемой памяти:  $O(c_1 |V| + c_2 |E|)$ , где  $c_1 = 2$ ,  $c_2 = 3$ .

Так как все вершины и связи хранятся в бинарных сбалансированных деревьях поиска, все оценки времени работы и объема занимаемой памяти имеют логарифмическую сложность. Для хранения всей структуры графа используется три базовые структуры данных: множество всех вершин, множество всех связей и структура, сопоставляющая каждой вершине множество ее инцидентных связей. Константа  $c_1$  получена за счет двойного хранения каждой вершины в первой и третьей структурах. Константа  $c_2$  получена за счет хранения каждой связи во второй структуре и двойного хранения каждой связи в третьей структуре данных.

В вышеописанной структуре данных в каждой вершине хранится вся необходимая информация для визуализации. Со связями дело обстоит несколько иначе: данные по изображению связи в качестве ломаной линии хранятся в отдельной структуре данных.

Введем следующие структуры данных. Одна структура данных будет содержать пару указателей на вершины с заданной операцией сравнения. Другая структура содержит координаты ломаной линии, которая выступает в качестве направляющей при прорисовке всех связей между двумя объектами. Связь будет визуализироваться с тем же количеством сегментов, идущих параллельно направляющей линии, за исключением двух: первого и последнего. Кроме того, во второй структуре также будет храниться дополнительная информация о различных параметрах сдвигов относительно направляющей линии.

Более того, экземпляр класса будет содержать в себе сбалансированное бинарное дерево поиска, ключами в котором будут являться указатели на связи. Это сделано для того, чтобы иметь быстрый доступ ко всем объектам, которые должны быть визуализированы относительно заданной направляющей кривой, так как для каждой связи не подразумевается хранение координат ее прорисовки. Все координаты будут вычисляться во время выполнения программы. Тогда можно создать сбалансированное бинарное дерево поиска, ключами в котором будут структуры, содержащие пары вершин, а значением будет структура с координатами ломаной линии.

Таким образом, мы сможем сократить объем хранимой памяти, тем самым эффективно решать

задачу хранения прорисовки связей и их быструю визуализацию.

При работе с большими графами критичными являются не только все без исключения базовые операции, но и объем требуемой памяти. Разработанные структуры данных эффективно сочетают в себе оба требования.

В программном комплексе предусмотрено несколько изобразительных соглашений для вершин: иконки, линии темы, таблицы. Отдельно стоит отметить способ представления вершин в виде линии темы. Линия темы – это вершина, которая представляется в виде горизонтального отрезка с иконкой, которая может перемещаться по всей длине отрезка. Если инцидентные линии темы вершины находятся в вертикальной полосе, ограниченной концами линии темы, то все связи между такими вершинами и линией темы будут вертикальными и перпендикулярно соединяться с линией темы. Остальные инцидентные линии темы вершины будут соединяться с ближайшими концами линии темы и не будут вертикальными. Положение концов линии темы и ее длина может меняться. Использование данного способа представления позволяет эффективно решать задачу визуализации для некоторых графов. Доступна визуализация множественных связей с изломами между вершинами, с возможностью задания ширины, цвета и типа линии. Для атрибутов также предусмотрено множество способов и настроек отображения.

Помимо богатой функциональности отображения различных элементов графа, разработанный модуль визуализации позволяет манипулировать большими графами, размеры которых достигают миллиона вершин и нескольких миллионов связей, за счет специально разработанных и оптимизированных алгоритмов машинной графики.

### 3. Средства автоматического размещения при визуализации графов

Модуль автоматического размещения объектов предназначен для целей визуального анализа графа и реализует графические схемы [6, 10] типа павлиний хвост, линия темы, круговое размещение, покомпонентное круговое размещение, а также размещение, основанное на алгоритме выделения сообществ в сети.

Размещение вершин и связей является нетривиальной задачей и в ручном режиме может требовать значительных временных затрат уже для графов

с количеством порядка десяти. Поэтому наряду с возможностью ручного размещения вершин и связей предоставляет несколько вариантов автоматического размещения элементов.

### 3.1. Случайное размещение объектов

Для заданной сети подбирается прямоугольная область нужного размера. Затем в эту область случайным образом помещаются все вершины и их связи. Распределение вершин в заданной области является равномерным.

### 3.2. Круговое размещение объектов

Узлы равномерно располагаются по одной окружности. Распределение по окружности производится на основе структуры связей. Минимизация пересечений в круговом размещении является NP-трудной задачей [7].

Пусть  $G(V, E)$  – граф, где  $V$  – множество вершин,  $E$  – множество ребер,  $n = |V|$ ,  $m = |E|$ . Используется двухфазный алгоритм [7] минимизации пересечений.

На первом этапе вершины добавляются в один их концов линейного размещения объектов. Пусть  $N(v) = \{u \in V : \{u, v\} \in E\}$  – множество смежных вершин для вершины  $v$ . Выбор первой добавляемой вершины определяется следующим образом:  $v_0 = \operatorname{argmax}_{v \in V}$  – вершина с максимальной степенью. Последовательность добавления остальных вершин определяется последовательностью обхода всех вершин начиная с вершины  $v_0$  алгоритмом обхода в ширину. После добавления всех вершин в линейное размещение оно естественным образом трансформируется в круговое путем соединения концов. Асимптотическая сложность данного этапа  $O(n)$ .

На втором этапе выполняется минимизация пересечения связей. Для каждой вершины  $v \in V$  рассматриваются все вершины  $u \in N(v)$  и выбирается та, при перестановке с которой количество пере-

сечений максимально уменьшится, после чего выполняется перестановка. Для заданной вершины процедура продолжается до тех пор, пока нельзя будет найти соседнюю вершину, перестановка с которой минимизирует количество пересечений. Асимптотическая сложность данной процедуры  $O(nm)$ . Экспериментальные результаты [7] показывают, что данную операцию достаточно сделать несколько раз для случайного набора вершин для достижения локального минимума функции количества пересечений.

### 3.3. Круговое покомпонентное размещение объектов

Сначала для исходной графа выполняется поиск всех компонент связности с помощью алгоритма обхода в ширину. В результате получено разбиение  $C$  графа  $G$  на непересекающиеся множества:

$$C = \{C_1, \dots, C_k \mid \bigcup_{i=1}^k C_i = V, \forall p, q: 1 \leq p, q \leq k, p \neq q, C_p \cap C_q = \emptyset\}.$$

Затем узлы каждой компоненты  $c_i$  располагаются по окружности на основе структуры их связей по алгоритму кругового размещения объектов. Центры окружностей размещаются на одной горизонтальной прямой. Пример представлен на рис. 2.

### 3.4. Размещение «одна линия темы»

Из сети выбирается вершина для представления в виде линии темы. Затем остальные вершины размещаются относительно заданной линии темы сверху и снизу последовательно друг за другом, то есть на каждом вертикальном уровне расположено не более двух вершин: одна сверху и одна снизу. Длина отрезка линии темы подбирается так, чтобы второстепенные объекты, соединенные с линией темы, соединились с ней вертикальными связями. Порядок расположения вторичных объектов можно конфигурировать различными способами. Пример размещения представлен на рис. 3.

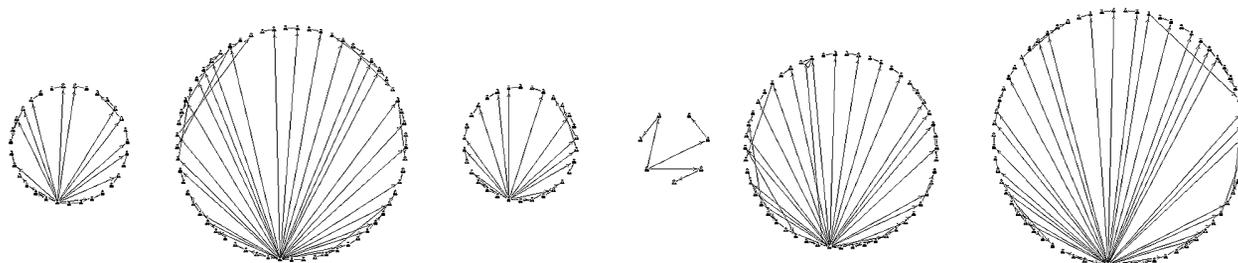


Рис. 2. Круговое покомпонентное размещение объектов

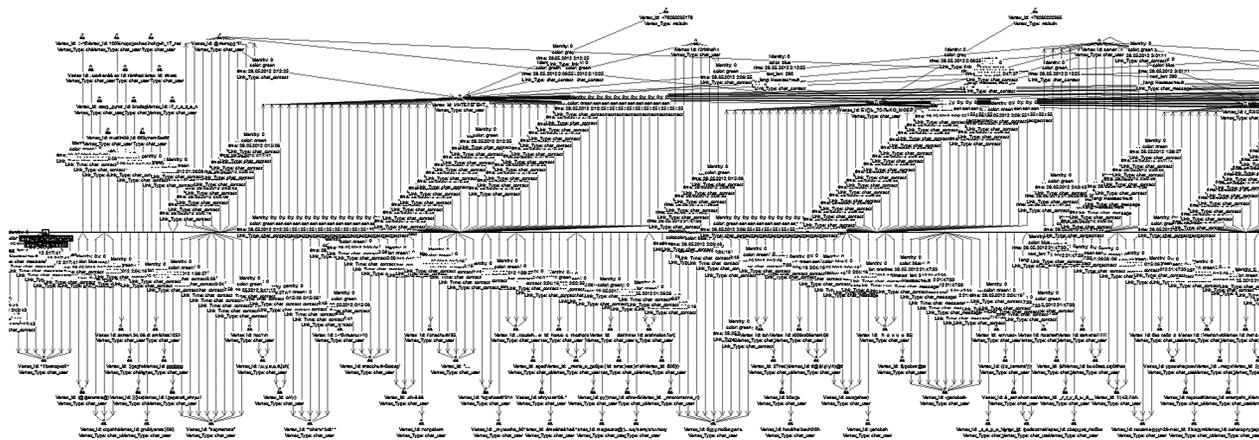


Рис. 3. Размещение «одна линия темы»

### 3.5. Размещение «две линии темы»

Пусть  $G(V, E)$  – граф, где  $V$  – множество вершин,  $E$  – множество ребер,  $n = |V|$ ,  $m = |E|$ .

Из сети выбираются два объекта для представления в виде линии темы  $V' = \{u_1, u_2\} \subseteq V$ . Назовем вершины из множества  $V \setminus V'$  вторичными.

Для визуализации графа  $G$  предлагается использовать многополосное размещение [2]. Проиллюстрируем предлагаемый способ на примере, когда выделенное множество  $V'$  содержит две вершины (рис. 4). Вершины выделенного множества  $u_1, u_2$  представляются в виде горизонтальных линий. Пространство между этими линиями разделено на три полосы: средняя полоса используется для размещения пометок вторичных вершин, а крайние полосы – для размещения пометок на ребрах. В общем случае, если выделенное множество  $V'$  содержит  $n$  вершин, то для размещения пометок вторичных вершин используется  $n+1$  полоса, а для размещения пометок на ребрах –  $2n$  полос соответственно. Для построения многополосного размещения предлагается использовать следующий алгоритм, состоящий из пяти шагов.

1. На выделенном множестве  $V'$  фиксируется некоторый порядок  $u_1, \dots, u_n$ , задающий относительное расположение горизонтальных линий, соответствующих вершинам выделенного множества.

2. Определяется порядок добавления ребер между вершинами выделенного множества.

3. Для каждой вторичной вершины  $v \in V \setminus V'$  выбирается полоса, в которой будет размещена пометка вершины  $v$ .

4. Определяется порядок добавления вторичных вершин  $v_1, \dots, v_m$ .

5. Путем добавления ребер между вершинами выделенного множества и последующего добавления вторичных вершин с инцидентными им ребрами производится построение многополосного размещения, гарантирующее отсутствие пересечений между пометками.

Опишем общие принципы работы алгоритма. Из графа выбираются два объекта для представления в виде линии темы. Линии темы размещаются на заданном расстоянии друг от друга. Затем остальные объекты размещаются относительно заданных линий тем сверху и снизу последовательно друг за другом. Длины отрезков линий тем подбираются так, чтобы второстепенные объекты, соединенные с линиями темы, соединялись с ними вертикальными связями. Стоит заметить, что все вершины, расположенные над первой линией темы соединены только с ней и между собой; все вершины, расположенные под второй линией темы соединены только с ней и между собой; между двумя линиями темы отображаются их связи и объекты, имеющие связи с обеими линиями темы. Порядок расположения вторичных объектов можно конфигурировать различными способами, пример на рис. 4.

### 3.6. Размещение, основанное на оценке связности

Данное размещение отличается от кругового покомпонентного размещения объектов тем, что между выделенными компонентами могут быть связи. Общая схема работы данного алгоритма размещения следующая: сначала задается порог на количе-

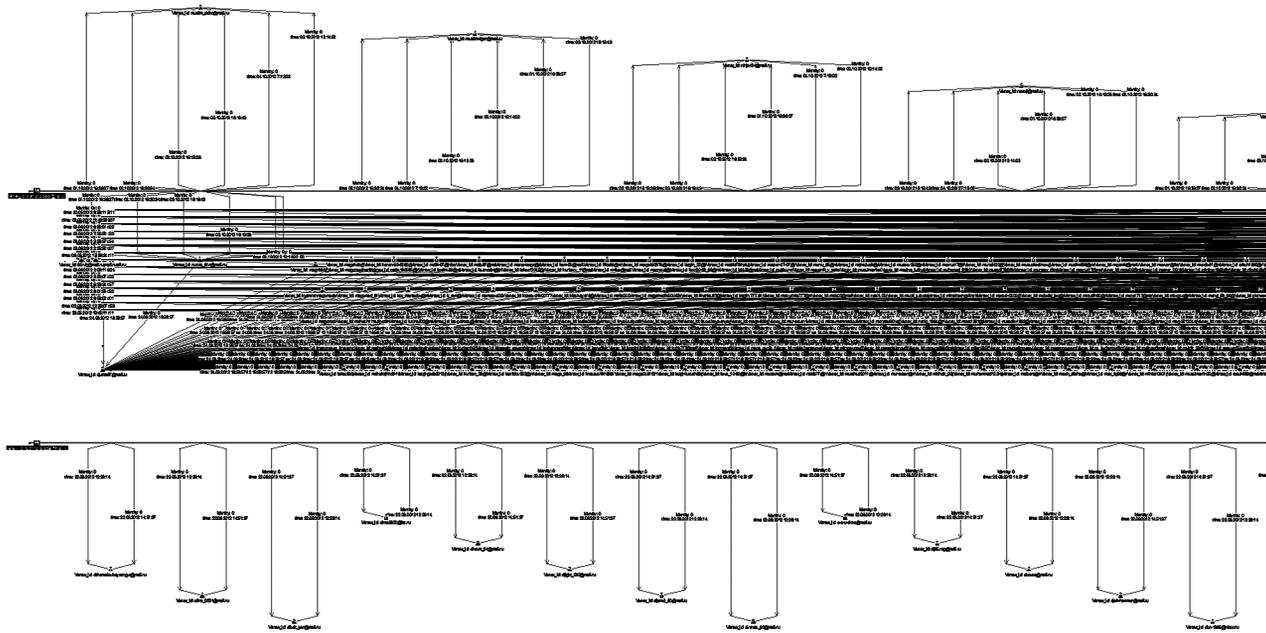


Рис. 4. Размещение «две линии темы»

ство связей между вершинами. Множество ребер  $E$  разбивается на два непересекающихся множества  $E = E_1 \cup E_2$  в соответствии с заданным порогом.  $E_1$  – множество ребер, удовлетворяющих заданному порогу,  $E_2$  – остальные. Далее применяется алгоритм выделения компонент связности на графе  $G' = (V, E_1)$ . В результате получено разбиение  $C$  вершин  $V$  исходного графа на непересекающиеся множества:

$$C = \{C_1, \dots, C_k \mid \bigcup_{i=1}^k C_i = V, \forall p, q: 1 \leq p, q \leq k, p \neq q, C_p \cap C_q = \emptyset\}.$$

Если количество связей между парой вершин ниже заданного порога, то они попадают в различные компоненты связности, иначе они попадают в одну компоненту связности. Затем узлы каждой компоненты  $C_i$  располагаются по окружности на основе структуры их связей по представленному выше алгоритму кругового размещения. Центры окружностей размещаются на одной горизонтальной прямой. На схеме также располагаются все связи графа, тем самым между вершинами различных компонент связности могут быть проведены связи.

### 3.7. Размещение, основанное на выделении сообществ

Реализован алгоритм автоматического выделения сообществ [3] с возможностью визуализации графа на основе полученного разбиения на сообщества.

Пример работы алгоритма выделения сообществ и применения размещения с использованием этого алгоритма представлен на рис. 5. Пусть  $L(M)$  – метрика качества полученного разбиения  $M$ .

$$L(M) = qH(Q) + \sum_{i=1}^m p_i H(P^i), \quad (1)$$

где  $M$  – разбиение сети на сообщества,  $m$  – количество сообществ,  $H(Q)$  – энтропия переходов между сообществами,  $H(P^i)$  – энтропия перемещения внутри сообщества  $i$ ,  $q$  – вероятность перехода между сообществами на каждом шаге случайного блуждания,  $p_i$  – вероятность остаться в сообществе  $i$ .

Метрика качества полученного разбиения может быть легко подсчитана для любого разбиения, обновление и пересчет этой метрики будет являться быстрой операцией.

С помощью алгоритма выделения сообществ все вершины разбиваются на непересекающиеся множества

$$C = \{C_1, \dots, C_k \mid \bigcup_{i=1}^k C_i = V, \forall p, q: 1 \leq p, q \leq k, p \neq q, C_p \cap C_q = \emptyset\}.$$

Каждое множество вершин представляет сообщество. Затем узлы каждого сообщества располагаются по окружности на основе структуры их связей. Центры окружностей размещаются по окружности. На схемы так же располагаются все связи графа, тем самым между вершинами различных сообществ могут быть проведены связи. Пример представлен на рис. 5.

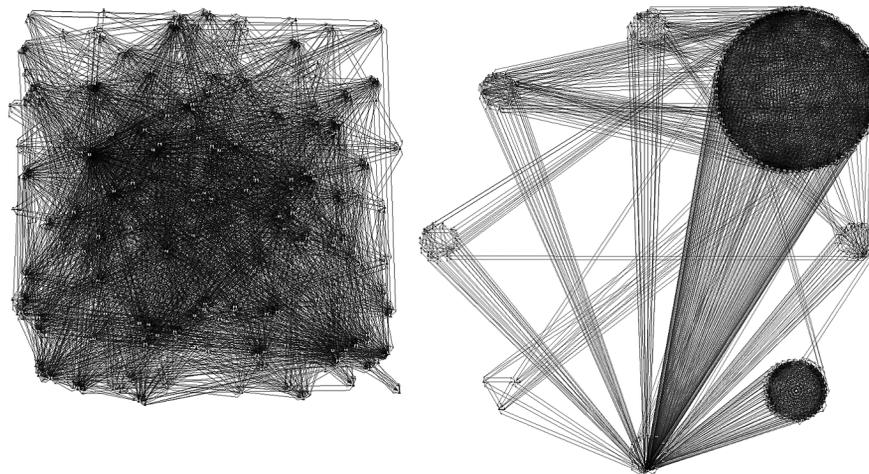


Рис. 5. Пример работы алгоритма выделения сообществ и размещения, основанного на этом алгоритме

### 3.8. Размещение «павлиний хвост»

Автоматическое размещение «павлиний хвост» базируется на методе физических аналогий. В качестве вершин выступают одинаково заряженные частицы, а в качестве связей – пружины. Отталкивающая сила вершин основывается на законе Кулона (2), описывающего силы взаимодействия между точечными электрическими:

$$\overline{F}_{12} = k \frac{q_1 q_2 \overline{r}_{12}}{r_{12}^2}, \quad (2)$$

где  $\overline{F}_{12}$  – сила, с которой заряд 1 действует на заряд 2,  $q_1, q_2$  – величины зарядов,  $\overline{r}_{12}$  – радиус-вектор (вектор, направленный от заряда 1 к заряду 2, и равный, по модулю, расстоянию между зарядами –  $r_{12}$ ),  $k$  – коэффициент пропорциональности.

Сила притяжения между связанными вершинами описывается законом Гука (3):

$$F = k \Delta l, \quad (3)$$

где  $F$  – сила, которой растягивают (сжимают) пружину,  $\Delta l$  – абсолютное удлинение (сжатие) пружины,  $k$  – коэффициент упругости.

Алгоритм авторазмещения является итерационным. Следовательно, алгоритм закончит свою работу либо в случае, когда было достигнуто максимальное количество итераций, либо когда динамическая система перешла в состояние равновесия.

### 4. Средства анализа графа

Модуль анализа графа предназначен для реализации алгоритмов и структур данных и позволяет про-

водить различные операции над графами. В данном модуле реализован алгоритм извлечения подграфов на основе поисков в ширину и на заданную в запросе глубину от заданной вершины, на основе используемых в запросах фильтров по атрибутам, на базе выделения кластеров. Предусмотрена возможность поиска объектов по заданным параметрам в соответствии с заданными алгоритмами. Кроме того, реализована функциональность поиска похожих вершин на основе структуры связей и значений атрибутов. Для более эффективного анализа графов реализованы алгоритмы объединения и пересечения графов. Также можно проанализировать кратчайшие пути между любой парой вершин в графе.

За счет связки с хранилищем графов удобно работать с огромными графами путем выгрузки подграфов по заданным фильтрам, объединения и пересечения графов. Стоит отметить, что за счет наличия модуля конвертации данных все графы хранятся в едином специализированном формате хранилища.

Аналитические инструменты позволяют проводить разнообразные исследования графа [11] и его структуры. Предусмотрен поиск и выделение вершин и связей по заданным атрибутам. Реализован подсчет метрик центральности для всех вершин и связей. Например, используется метрика связи – центральность по посредничеству (4) (betweenness centrality): количество кратчайших путей между всеми парами вершин, проходящих через заданную связь.

$$betw(l) = \sum_{s \neq l, s \in V} \frac{\sigma_{st}(l)}{\sigma_{st}}, \quad (4)$$

где  $l$  – связь, для которой считается метрика,  $V$  – множество вершин графа,  $\sigma_{st}$  – количество кратчайших путей между вершинами  $s$  и  $t$ ,  $\sigma_{st}(l)$  – коли-

чество кратчайших путей между вершинами  $s$  и  $t$ , проходящих через связь  $l$ .

При работе с большими графами для анализа структуры реализован алгоритм оценки связности, который позволяет для заданного порога по числу связей между парой вершин разбить граф на компоненты связности, выполнить размещение объектов с учетом полученных компонент и показать списки всех вершин для каждой полученной компоненты.

### Заключение

Описанный программный комплекс реализован на языке C++. Реализация визуализации выполнена с использованием графической библиотеки

OpenGL. Интерфейс реализован с использованием библиотеки Qt. Все используемые средства реализации являются кроссплатформенными. Тем самым программный комплекс способен функционировать в различных операционных системах.

Предлагаемый в данной статье программный комплекс предоставляет широкую функциональность при работе с графами больших размеров. В данном программном обеспечении реализованы средства представления графов, позволяющие проводить анализ графа взаимодействующих объектов визуальными средствами после алгоритмической обработки. Данное средство предназначено для широкого круга аналитиков в областях информационной безопасности, маркетинга, социологических исследований. ■

### Литература

1. Батура Т.В. Методы анализа компьютерных социальных сетей // Вестник НГУ, серия «Информационные технологии». 2012. Т. 10, вып. 4. С. 13-28.
2. Борисенко В.В., Лахно А.П., Чеповский А.М. Специальное представление графов и визуализация семантических сетей // Фундаментальная и прикладная математика. 2010. Т. 16, вып. 8. С. 27-35.
3. Коломейченко М.И., Чеповский А.А., Чеповский А.М. Алгоритм выделения сообществ в социальных сетях // Фундаментальная и прикладная математика. 2014. Т. 19, №1. С. 21-32.
4. Поляков И.В., Чеповский А.А., Чеповский А.М. Хранение и обработка графа социальных сетей // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. Т. 11, № 4. С. 77-83.
5. Чураков А.Н. Анализ социальных сетей // СоцИс. 2001. №1. С. 109-121.
6. Battista G., Tamassia R., Tollis I. Graph drawing: Algorithms for the visualization of graphs. Springer, 1999.
7. Baur M., Brandes U. Crossing reduction in circular layouts // Proceedings of the 30th International Conference on Graph-Theoretic Concepts in Computer Science (WG 2004). Bonn, Germany, 21-23 June 2004. P. 332-343.
8. CrimeLink. URL: <http://www.pciusa.us/Crimelink.aspx> (дата обращения: 05.04.2014).
9. i2 Analyst's Notebook. URL: <http://www-03.ibm.com/software/products/ru/analysts-notebook> (дата обращения: 05.04.2014).
10. Kaufmann M., Wagner D. Drawing Graphs. Springer, 2001.
11. Newman M.E. Networks: An Introduction. Oxford, UK: Oxford University Press, 2010.
12. Sentinel Visualizer. URL: <http://www.fmsag.com/Products/SentinelVisualizer> (дата обращения: 05.04.2014).
13. Tom Sawyer. URL: <http://www.tomsawyer.com> (дата обращения: 05.04.2014).
14. XAnalys Link Explorer. URL: <http://www.xanalys.com/solutions/linkexplorer.html> (дата обращения: 05.04.2014).

# HUGE GRAPH VISUALIZATION AND ANALYSIS

## **Maxim KOLOMEYCHENKO**

Graduate of MSc Program, National Research University Higher School of Economics

Address: 20, Myasnitskaya street, Moscow, 101000, Russian Federation

E-mail: maxim.kolomeychenko@mail.ru

## **Andrey CHEPOVSKIY**

Associate Professor, Department of Information Security Management, Faculty of Business Informatics, National Research University Higher School of Economics;

Professor, Department of Applied Mathematics and Systems Modeling, Institute of Communications and Media Business, Moscow State University of Printing Arts

Address: 2a, Pryanishnikova street, Moscow, 127550, Russian Federation

E-mail: achepovskiy@hse.ru

The problem of huge graph visualization arises in various fields of sociology and marketing. The relevance of this work is determined by the need for the software package for analysis and visualization of such graphs. This paper introduces summary analysis of several software products and highlights their weaknesses: lack of cross-platform and specialized graph warehouses, inability to deal with huge graphs.

This paper presents a detailed description of the overall architecture of the software developed and each of its modules separately, as well as a procedure for communication between core modules. A special graph warehouse capable to process graphs having up to 100 million vertices and billions of links is used to store graphs. A description of main data warehousing principles is also introduced. The use of a proprietary file system ensures the absence of additional system calls when working with the warehouse and the lack of a complex addressing system and any excessive mechanisms, that enables to avoid any extra overhead costs associated with data warehousing. Furthermore, there is a description of a methodology for data visualization module, used data structures and computer graphics algorithms, that enable to handle graphs comprising up to millions of vertices on real time basis.

It is worth noting a wide range of algorithms for auto layout of graphs such as random layout, circular layout, circular component layout, peacock tail layout, one or two themes layout, layout based on community detection and estimation of cohesion. This paper presents a detailed description of each of the layout mentioned above.

Special attention should be paid to the developed methods of graph analysis. Algorithms have been designed to detect communities in social networks, to evaluate graph cohesion, to find the shortest paths between any pair of vertices, graphs union and intersection, etc.

A key feature of all algorithms presented in this paper is their ability to handle huge graphs.

**Key words:** graph, graph analysis, graph visualization, graph storage, software.

### References

1. Batura T. (2012) Metody analiza kompjuternyh socialnyh setej [Methods of social networks analysis]. *Vestnik NGU, serija «Informacionnye tehnologii»* [Information Technologies], no. 10 (4), pp. 13-28. (in Russian)
2. Borisenko V.V., Lahno A.P., Chepovskiy A.M. (2010) Specialnoe predstavlenie grafov i vizualizacija semanticheskikh setej [Special graph representation and visualization of semantic networks]. *Fundamentalnaya i prikladnaya matematika* [Fundamental and Applied Mathematics], vol. 16, no. 8, pp. 27-35. (in Russian)
3. Kolomejchenko M.I., Chepovskiy A.A., Chepovskiy A.M. (2014) Algoritm vydelenija soobshhestv v socialnyh setjah [The algorithm for detecting communities in social networks]. *Fundamentalnaya i prikladnaya matematika* [Fundamental and Applied Mathematics], vol. 19, no. 1, pp. 21-32. (in Russian)
4. Poljakov I.V., Chepovskiy A.A., Chepovskiy A.M. (2013) Hranenie i obrabotka grafa socialnyh setej [Social networks storing and processing]. *Vestnik NGU, serija «Informacionnye tehnologii»* [Information Technologies], vol. 11, no. 4, pp. 77-83. (in Russian)
5. Churakov A.N. (2001) Analiz sotsialnykh setej [Social network analysis]. *SotsIs*, no. 1, pp. 109-121. (in Russian)
6. Battista G., Tamassia R., Tollis I. (1999) *Graph drawing: Algorithms for the visualization of graphs*. Springer.
7. Baur M., Brandes U. (2004) Crossing reduction in circular layouts. Proceedings of the *30th International Conference on Graph-Theoretic Concepts in Computer Science (WG 2004)*, WG 2004, Bonn, Germany, 21-23 June, 2004, pp. 332-343.
8. CrimeLink. Available at: <http://www.pciusa.us/Crimelink.aspx> (accessed: 05.04.2014).
9. i2 Analyst's Notebook. Available at: <http://www-03.ibm.com/software/products/ru/analysts-notebook> (accessed: 05.04.2014).
10. Kaufmann M., Wagner D. (2001) *Drawing Graphs*. Springer.
11. Newman M.E. (2010) *Networks: An Introduction*. Oxford, UK: Oxford University Press.
12. Sentinel Visualizer. Available at: <http://www.fmsag.com/Products/SentinelVisualizer> (accessed: 05.04.2014).
13. Tom Sawyer. Available at: <http://www.tomsawyer.com> (accessed: 05.04.2014).
14. XAnalys Link Explorer. Available at: <http://www.xanalys.com/solutions/linkexplorer.html> (accessed: 05.04.2014).

# АВТОМАТИЧЕСКИЙ ПОИСК ОПОРНЫХ ЭЛЕМЕНТОВ НА ДОКУМЕНТАХ ПОЛУЖЕСТКОЙ СТРУКТУРЫ

**М.О. ЛАНИН**

аспирант кафедры распознавания изображений и обработки текста,  
факультет инноваций и высоких технологий, Московский физико-технический  
институт (государственный университет); программист, ООО «Аби Продакшн»

Адрес: 141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9

E-mail: mike.lanin@gmail.com

Статья посвящена проблеме потокового извлечения данных из документов полужесткой структуры, для которых слабо применимы методы сплошного оптического распознавания символов. Для облегчения процесса создания структурных описаний таких документов широко используются методы машинного обучения. Тем не менее, существующие решения по-прежнему достаточно сложны для конечного пользователя, поскольку требуют ручного описания элементов структуры документа, не имеющих прямого отношения к извлекаемым данным.

В работе рассматривается возможный подход к описанию изображений документов переменной структуры, используемый в системе потокового ввода ABBYY FlexiCapture, а также метод автоматического построения такой структурной модели по разметке всех элементов структуры. Подробно описывается алгоритм автоматического поиска опорных элементов по пользовательской разметке извлекаемых данных, позволяющий значительно облегчить с точки зрения пользователя процесс создания структурной модели документа ABBYY FlexiCapture. Интеграция описанной технологии обучения на этапе верификации извлекаемых данных позволяет инкрементально улучшать структурную модель документа, при этом все, что требуется от конечного пользователя — исправлять неверно найденные в процессе ввода регионы извлекаемых полей. Также в статье описан метод и результат оценки эффективности предложенного подхода. Описанный способ поиска опорных элементов показал свою применимость на реальных платежных документах ряда немецких поставщиков: 89,3% счетов могут быть обработаны без ошибок при минимальном участии пользователя, при этом верно извлечены данные из 97,8% полей.

**Ключевые слова:** система потокового ввода, документы полужесткой структуры, структурное описание изображения документа, опорные элементы, реперы, поля, разметка полей, методы машинного обучения, частотный словарь.

## 1. Введение

По мере все большего распространения распределенных информационных систем и систем электронного документооборота набирают популярность и системы потокового ввода данных. На вход таких систем попадает большой массив документов различной природы, поэтому обработка всего объема данных вручную затруднительна, затратна по времени и практически невозможна. Построение эффективной системы потоко-

вого ввода документов связано с решением целого ряда задач, таких как распознавание текста, классификация изображений, маршрутизация и извлечение данных из документов.

В данной статье внимание уделено проблеме автоматического извлечения данных из изображений документов. Даже документы одного типа зачастую имеют полужесткую структуру, то есть расположение однотипных данных на них варьируется от одного экземпляра к другому. Использование

трафаретных форм с явным указанием регионов извлекаемых данных и дальнейшим распознаванием текста методами оптического распознавания символов не позволяет корректно извлекать информацию из таких документов. Для извлечения информации из документов нежесткой логической и графической структуры используется специализированная структурная модель документа [1; 2]. Большую трудность в использовании таких моделей представляет необходимость прямого указания геометрических отношений между элементами документа [3; 4], создания тематических словарей [2; 3], а также ручного описания некоторых элементов структуры [2; 4]. Все это обуславливает высокие требования к квалификации создателей структурных описаний и большие временные затраты на создание эффективных описаний даже для документов простой структуры.

Для упрощения процесса создания структурных описаний документов используются методы машинного обучения. Показал свою эффективность метод автоматического построения структурного описания документа с использованием пользовательской разметки всех элементов [5]. Однако, этот подход хоть и значительно упрощает процесс создания структурных описаний, но не решает проблему в целом. От пользователя по-прежнему требуется вручную указывать набор статических элементов (подписи, заголовки и т.д.), с этим связаны определенные трудности. Во-первых, пользователь не всегда может хорошо определить, является ли строка «хорошим» опорным элементом (репером), надежно локализирующим поле. Во-вторых, возникает ряд проблем, если предполагается, что процесс создания структурного описания может быть распределен по времени или пространству, так как набор размеченных реперных элементов должен совпадать для всех документов обучающей выборки. Следующим шагом в упрощении процесса создания структурного описания является автоматизация поиска опорных элементов документа на основе разметки извлекаемых данных (полей). Такой подход до минимума снижает требовательность к квалификации пользователя, поскольку требует лишь указания положения полей на изображении документа, а также позволяет внедрить систему обучения на этапе верификации извлекаемых данных и инкрементально улучшать структурное описание в процессе ввода.

В работе рассматривается подход к автоматическому поиску реперных элементов для дальнейше-

го построения структурного описания в рамках модели, используемой в системе ABBYY FlexiCapture. Применимость предложенного метода была проверена на тестовом пакете из 622 реальных немецких счетов от 10 поставщиков. Метод показал свою эффективность, как часть процесса инкрементального обучения структурного описания в процессе ввода.

## 2. Модель структурного описания документов

Методы автоматического поиска опорных элементов рассматриваются в контексте модели структурного описания, используемой в системе потокового ввода ABBYY FlexiCapture. Структурное описание документа представляет собой дерево типизированных элементов, определяющих формат и содержимое некоторых частей документа. Всего доступно 18 типов элементов для различных типов данных, таких как статический текст, дата, денежная сумма, число, телефон, регулярное выражение и т.д. Для каждого из элементов задается набор свойств, определяющих как содержимое всего документа (обязательные и запрещенные элементы, минимальное и максимальное количество подэлементов в группе, количество повторений и т.д.), так и содержимое самих элементов (варианты текста, минимальное и максимальное значение, формат, количество символов, количество строк и т.д.). Кроме того, для каждого элемента могут быть заданы ограничения области поиска – набор полуплоскостей, в которых элемент может быть найден. Ограничения области поиска могут быть заданы относительно границ изображения либо относительно расположения других, найденных ранее, элементов.

На основе заданных свойств и областей поиска определяется качество гипотезы о том, что распознанное слово или множество слов является соответствующим элементом. Нарушение каждого из правил приводит к наложению определенного штрафа, размер которого зависит от степени нарушения и пользовательских настроек. В процессе анализа документа строится дерево гипотез: для каждой из гипотез расположения текущего элемента выдвигается множество гипотез расположения следующих элементов. Качество цепочки гипотез рассчитывается как произведение качеств каждой из гипотез этой цепочки. Результатом анализа является цепочка гипотез наилучшего качества, если

качество такой цепочки превосходит предельное минимальное значение; в противном случае считается, что документ не соответствует заданному структурному описанию. В общем случае логика построения структурного описания заключается в определении статического «каркаса» документа и последующей локализации извлекаемых данных относительно его.

### 3. Метод построения структурного описания по пользовательской разметке

Подробное описание метода автоматического построения структурного описания документа по пользовательской разметке выходит за рамки данной статьи, однако для общего понимания рассмотрим краткое описание используемого алгоритма. Для задачи построения структурного описания не рассматривается проблема классификации документов: считается, что структурное описание создается для документов схожей структуры (счета выделенного поставщика, анкеты определенного клиента и т.д.). В случае, когда на вход поступают документы различной структуры, подразумевается использование внешнего классификатора и нескольких структурных описаний для каждого из типов обрабатываемых документов.

С учетом вышеуказанных предположений и описанной ранее модели структурного описания, задача автоматического создания структурного описания по пользовательской разметке всех элементов сводится к двум подзадачам: определение типа и настройка свойств каждого из элементов и построение геометрических отношений между элементами. Набор свойств элемента зависит от его типа, и алгоритм их настройки не будет рассматриваться подробно. В общем случае, свойства задаются как наиболее строгие разумные ограничения, включающие все возможные варианты обучающей выборки.

Для построения геометрических ограничений области поиска для каждого документа строится матрица отношений между элементами типа «вышележе» и «левее-правее», после чего результирующая матрица отношений получается как пересечение всего множества матриц. Отступы ограничений от границ элементов также задаются как наиболее строгие ограничения, включающие все возможные варианты взаимного расположения элементов на документах набора обучения.

### 4. Поиск опорных элементов с использованием пользовательской разметки полей

Вкратце, поиск опорных элементов состоит из следующих этапов:

1. построение частотного словаря для документов обучающей выборки;
2. выбор слов-кандидатов в реперные элементы по частотности;
3. слияние геометрически близких слов-кандидатов в строки для каждого из документов;
4. сопоставление реперов на документах обучающей выборки.

Для поиска опорных элементов для всех документов обучающей выборки строится частотный словарь с учетом возможных ошибок распознавания: слова считаются вариантами одного слова, если сводятся друг к другу не превосходящим порогового значения количеством элементарных операций, к которым относятся вставка, удаление и замена символа.

На основе частотных словарей выбираются слова-кандидаты, при этом для оценки качества слова  $w$  используется мера

$$q(w) = \frac{\sum_i q(d_i, w)}{|D|},$$

где  $|D|$  – количество документов в наборе обучения, а  $q(d_i, w)$  – качество слова  $w$  на документе  $d_i$ , равное  $\max_j(q(d_i, w_j))$  – максимальному качеству конкретного экземпляра слова на документе. Качество каждого конкретного экземпляра слова вычисляется в соответствии с форматом содержимого, форматированием, геометрическими размерами и расположением текста. Так, бонус получают слова форматно похожие на заголовки, URL, подписи и т.п. статические элементы документа. Слова, входящие в мелкий и сплошной текст, получают меньшее качество. Также штрафуются слова по формату похожие на численные значения, даты, пунктуаторы и т.п.

Дополнительно отсекаются слова с малой частотностью, по мере

$$df(w) = \log \left( 1 + \frac{|d_i \ni w|}{|D|} \right),$$

где  $|d_i \ni w|$  – количество документов обучающей выборки, в которых встречается слово  $w$ , а  $|D|$  – общее количество документов в наборе обучения.

На этом этапе не используется мера для отсекаания слов, часто встречающихся на одном экземпляре документа, поскольку частоупотребительное в контексте документа слово может быть частью менее частотной строки, являющейся хорошим опорным элементом. Например, частотные на платежных документах слова «invoice» и «date» образуют надежный репер-подпись к полю «invoice date».

На следующем этапе геометрически близкие слова-кандидаты сливаются в строку на каждом из документов набора обучения. Список опорных элементов получается путем сопоставления полученных строк-кандидатов для каждого из документов. Начальный список реперов формируется на основе строк-кандидатов, найденных на первом документе, дальнейший процесс сопоставления происходит итеративно следующим образом:

1. Для каждого следующего документа строится двудольный граф, вершинами которого являются строки-кандидаты и имеющиеся на данном этапе реперы, соответственно. Вес каждого ребра для репера  $r$  и строки-кандидата  $s$  рассчитывается по разметке множества полей  $F$  как

$$W(r, s) = T(r, s) \cdot (\max_i L(r, s, f_i) - k \frac{\sum_{j=1}^N L(r, s, f_j)}{|F|}),$$

где  $T(r, s)$  – степень совпадения текстового содержания  $r$  и  $s$ ,  $L(r, s, f_i)$  – степень совпадения относительного расположения поля  $f_i$  относительно  $r$  на коллекции обработанных документов и относительно  $s$  на текущем документе,  $|F|$  – общее количество полей, а  $k$  – константа.

2. Из графа удаляются все ребра с качеством ниже порогового, после чего ищется максимальное паросочетание, определяющее принадлежность строки кандидата к соответствующему реперу. Строки, не имеющие парного репера, добавляются как новый реперный элемент.

3. Для всех реперных элементов обновляются данные об относительном расположении полей с учетом последнего обработанного документа.

На последнем этапе, после окончания процесса сопоставления, из полученного списка реперных элементов удаляются элементы, представленные на малом относительном количестве документов обучающей выборки.

## 5. Методика и результат тестирования

Оценка эффективности метода автоматического поиска реперных элементов сама по себе за-

труднительна, поскольку, во-первых, не всегда можно объективно определить, насколько качественным репером является конкретная строка, во-вторых, зачастую для построения качественного структурного описания документа не обязательно использовать все «надежные» статические элементы, а достаточно ограничиться каким-то разумным минимумом. В связи с этим характеристики полноты и точности относительно некоторой эталонной разметки реперов не являются приемлемыми показателями качества вне контекста дальнейшего использования найденных опорных элементов.

Предложенный подход тестируется в связке с описанным ранее методом автоматического создания структурных описаний, как первый шаг инкрементального обучения структурного описания в процессе ввода: сперва по пользовательской разметке полей определяется расположение опорных элементов, затем объединение разметки реперов и полей используется для создания очередной версии структурного описания. Эффективность такого подхода сравнивается с автоматическим созданием структурного описания с использованием вручную указанных опорных элементов: для документов обучающей выборки оператор указывает расположение подписей к извлекаемым данным, а также без использования опорных элементов, исключительно по положению полей.

В общем виде сценарий тестирования выглядит следующим образом:

1. Для начала обработки документов, пользователь вручную указывает расположение извлекаемых данных на первых трех изображениях документов данного типа.

2. По окончании разметки автоматически создается первая версия структурного описания, используемого в дальнейшем для извлечения данных из документов.

3. Если в процессе потоковой обработки какие-то поля нашлись неверно, пользователь вручную исправляет расположение полей.

4. В случае, когда ошибки обработки не связаны с дефектами изображения (проблемы распознавания, перекосы и т.д.), пользователь добавляет проблемный экземпляр документа в набор обучения.

5. На основе пользовательской разметки полей по всему набору обучения создается очередная версия структурного описания, которая используется для дальнейшей обработки документов.

Для проверки применимости предложенного подхода использовался пакет, содержащий 622 реальных счета от 10 немецких поставщиков. Таким образом, в процессе ввода участвовало 592 документа, поскольку первые три изображения документов каждого из поставщиков использовались для создания исходной версии структурного описания.

Таблица 1.

**Результаты обработки документов при обучении без использования реперов**

Количество ошибок	Количество документов	Доля документов
0	0	0%
1	0	0%
2	6	1,0%
3	62	10,5%
4	58	9,8%
5	134	22,6%
6	260	43,9%
7	72	12,2%

Результаты эксперимента, приведенные в табл. 1, показывают, что метод обучения без использования статических элементов не пригоден для потокового ввода. Так, ни один документ не был введен без ошибок, менее 3 ошибок было допущено всего на 1% документов, а для более, чем половины документов потока выделено не более одного поля.

Таблица 2.

**Результаты извлечения данных полей при обучении без использования реперов**

Имя поля	Количество полей в пакете	Корректно найдено полей	Доля найденных полей
Получатель	592	161	27,2%
Адрес получателя	590	64	10,8%
Номер счета	574	191	33,2%
Дата	592	372	62,8%
Сумма без налога	580	10	1,7%
Сумма налога	492	57	11,6%
Сумма	588	28	4,8%

Из табл. 2 видно, что более-менее приемлемый результат извлечения данных без использования опорных элементов достигается только для достаточно жестко позиционированных полей выделенного формата (дата). Жестко располагающиеся поля более свободного формата выделяются сильно хуже (получатель, номер счета), а качество выделения полей с плавающей позицией (сумма счета) или переменного размера (адрес получателя) нельзя назвать хоть сколько-нибудь приемлемым.

Таблица 3.

**Сравнение результатов обработки документов при автоматическом поиске и ручном указании опорных элементов**

Количество ошибок	Количество документов		Доля документов	
	авто	ручной	Авто	Ручной
0	529	418	89,3%	70,6%
1	46	122	8,6%	20,6%
2	9	33	1,5%	5,6%
3	3	10	0,5%	1,7%
4	2	5	0,3%	0,8%
5	0	1	0%	0,1%
6	3	3	0,5%	0,5%

Таблица 4.

**Сравнение результатов извлечения данных полей при автоматическом поиске и ручном указании опорных элементов**

Имя поля	Количество полей в пакете	Корректно найдено полей		Найдено лишних полей	
		авто	ручной	авто	ручной
Получатель	592	577	555	0	0
Адрес получателя	590	571	467	2	0
Номер счета	574	551	536	4	4
Дата	592	585	572	0	0
Сумма без налога	580	570	570	8	12
Сумма налога	492	485	478	2	2
Сумма	588	580	575	0	0

бл. 3 и 4 приведено сравнение результатов ввода с инкрементальным обучением структурного описания при использовании автоматически найденных («авто») и указанных оператором вручную («ручной») опорных элементов. Описанный метод показал свою эффективность: 89,3% документов были обработаны без ошибок, а ошибка более чем в два поля была допущена менее, чем на 1,5% документов, при этом успешно было найдено 3919 из 4008 полей (97,8%). Более того, в контексте описанного инкрементального обучения структурной модели, предложенный подход оказался эффективнее, чем ручное указание оператором позиций подписей к извлекаемым полям.

### 6. Выводы

Метод показал свою эффективность и позволил минимизировать участие пользователя в процессе

создания структурного описания. Результаты тестового эксперимента показали, что пользователь может получить структурное описание достаточно высокого качества, просто указывая расположение извлекаемых данных. Тем не менее, по-прежнему требуется поддержка пользователя для фильтрации дефектных изображений, что требует определенной, хоть и не высокой, квалификации. В идеале, система должна уметь автоматически принимать решение, добавлять ли очередное изображение в обучающую выборку.

Также не стоит забывать, что при разработке метода использовалось достаточно сильное предположение о том, что документы различной природы и структуры могут быть наверняка разделены внешним классификатором. Однако, само по себе создание такого классификатора — достаточно сложная и трудоемкая задача, требующая отдельного исследования. ■

### Литература

1. Зуев К.А. Система идентификации структуры печатных документов. Дисс. канд. тех. наук. М.: МГУЛ, 1999.
2. Hamza H., Belaid A., Belaid Y., Chaudhuri B. An end-to-end administrative document analysis system // Proceedings of the Eighth International Workshop on Document Analysis Systems (DAS 2008), Nara, Japan, September 16-19, 2008. P. 175-182.
3. Ishitani Y. Model-based information extraction method tolerant of OCR errors for document images // Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR), Seattle, Washington, USA, September 10-13, 2001. P. 908-915.
4. Peanho C., Stagni H., da Silva F. Semantic information extraction from scanned images of complex documents // Applied Intelligence. December 2012. Vol. 37, No. 5. P. 543-557.
5. Голубев С.В. Распознавание структурированных документов на основе машинного обучения // Бизнес-информатика. 2011. № 2. С. 48-55.

# AUTOMATIC DETECTION OF REFERENCE ELEMENTS ON SEMI-STRUCTURED DOCUMENT IMAGES

**Mikhail LANIN**

Post-graduate student, Department of Images Recognition and Text Processing, Faculty of Innovations and High Technologies, Moscow Institute of Physics and Technology (State University); Software engineer, ABBYY Production

Address: 9, Institutskiy per., Dolgoprudny, Moscow Region, 141700, Russian Federation

E-mail: mike.lanin@gmail.com

*The paper deals with automatic data extraction from semi-structured documents. The through optical character recognition methods are slightly applicable for this kind of input. To simplify the process to create structural descriptions of such documents machine learning methods are widely used, however, current solutions are still complicated for end-users, because these require manual description of document structure elements, which are not directly relevant to date to be extracted.*

*The article presents a possible approach to describe variable structure document images used in document data capture system called ABBYY FlexiCapture and a method of automatic model creation based on layout of all structure elements. The paper provides a detailed description of an algorithm for automatic detection of reference elements based on user layout of data to be extracted that enables to facilitate dramatically the process of building of a structured model of an ABBYY FlexiCapture document from the user perspective. Integration of this technology at the data extraction validation stage enables to incrementally improve the structural model of a document, as it requires a user only to correct localization of wrongly found data being extracted. Finally, the paper describes a method to assess robustness of the proposed approach and test results. The described method involving detection of reference elements has shown its effectiveness in processing actual payment documents of a number of German suppliers: 89.3% of invoiced can be treated with no faults with minimum user intervention; furthermore, the data had been extracted correctly from 97.8% of fields.*

**Key words:** data capture, semi-structured documents, structural description of document, reference elements, reference points, fields, fields layout, machine learning, frequency list.

## References

1. Zuev K.A. (1999) *Sistema identifikatsii struktury pechatnykh dokumentov* (diss. kand. tekhn. nauk) [Printed documents structure identification system (PhD Thesis)]. Moscow: MGUL. (in Russian)
2. Hamza H., Belaid A., Belaid Y., Chaudhuri B. (2008) An end-to-end administrative document analysis system. Proceedings of Proceedings of the *Eighth International Workshop on Document Analysis Systems (DAS 2008)*, Nara, Japan, September 16-19, 2008. P. 175-182.
3. Ishitani Y. (2001) Model-based information extraction method tolerant of OCR errors for document images. Proceedings of the *Sixth International Conference on Document Analysis and Recognition (ICDAR)*, Seattle, Washington, USA, September 10-13, 2001, pp. 908-915.
4. Peanho C., Stagni H., da Silva F. (2012) Semantic information extraction from scanned images of complex documents. *Applied Intelligence*, vol. 37, no. 5, pp. 543-557.
5. Golubev S.V. (2011) Raspoznavanie strukturovannykh dokumentov na osnove mashinnogo obucheniya [Structured documents recognition based on machine learning]. *Business Informatics*, no. 2 (16), pp. 48-55. (in Russian)

# АЛГОРИТМ ФОРМИРОВАНИЯ МНОГОКРИТЕРИАЛЬНОЙ СТРАТИФИКАЦИИ

**М.А. ОРЛОВ**

аспирант кафедры анализа данных и искусственного интеллекта,  
факультет компьютерных наук, Национальный исследовательский  
университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: ortmian@mail.ru

Данная работа развивает подход к проблеме многокритериального ранжирования, называемый нами многокритериальной стратификацией. При таком подходе представляет интерес не столько полное упорядочивание вариантов, сколько разбиение множества вариантов на заданное число классов, упорядоченных по предпочтению. Ранжирование получается путем линейной свертки критериев с весами. При этом веса определяются, исходя из предположения о наличии некоторой структуры в данных, такой что можно выделить «параллельные» слои – страты.

В работе [6] авторы сформулировали задачу формирования оптимальной стратификации, как задачу оптимизации некоторой целевой функции относительно весов критериев, однако, предложенный в этой работе алгоритм решения получаемой задачи, использующий случайный поиск, оказался неэффективным по сравнению с другими методами стратификации.

В данной работе предлагается новый алгоритм оптимизации целевой функции многокритериальной стратификации на основе квадратичного программирования. Для всестороннего экспериментального исследования качества работы алгоритма предлагается усовершенствованная модель генерации искусственных стратифицированных данных. Новая модель генерации страт имеет больше параметров для настройки и позволяет гибко задавать геометрию страт: ориентацию, толщину, размах и интенсивность, что лучше учитывает структуру реальных данных. Предлагаемый алгоритм экспериментально сравнивается с существующими методами стратификации на искусственных данных, и показывается его преимущество в большинстве рассмотренных случаев. Рассматриваются два примера реальных данных – библиометрические показатели 118 научных журналов и характеристики публикационной активности 102 стран. На этих данных новый алгоритм приводит к хорошо интерпретируемым и адекватным результатам. Также оказалось, что на этих данных построенное алгоритмом многокритериальное разбиение наиболее согласовано с разбиениями, построенными по отдельно взятым критериям.

**Ключевые слова:** стратификация, многокритериальное ранжирование, взвешенная сумма, квадратичное программирование, оптимизация, библиометрия.

## Введение

В работе [6] рассматривается проблема стратификации как поиск весов критериев таких, что элементы страт образуют компактные группы на оси обобщенного критерия. Для решения возникающей оптимизационной задачи определения оптимальных весовых коэффициен-

тов критериев в этой работе использовался метод случайного поиска, имитирующий эволюцию. Предложенный метод стратификации экспериментально сравнивался с рядом других алгоритмов и, в общем и целом, уступил некоторым из них. В данной работе предлагается использовать более точный метод квадратичного программирования вместо случайного поиска. В такой модификации

метод стратификации оказывается вполне конкурентоспособным в вычислительных экспериментах. Применение метода к реальным данным приводит к хорошо интерпретируемым весовым коэффициентам критериев, а получаемые страты оказываются достаточно адекватны. Для проведения контролируемого вычислительного эксперимента на искусственных данных в настоящей работе предлагается усовершенствованная параметрическая модель генерации линейных страт. Этот механизм генерации позволяет в зависимости от параметров гибко задавать структуру страт, и тем самым моделировать многие ситуации, возникающие в задачах принятия решений.

Дальнейшее содержание включает четыре раздела. В первом разделе приводится краткий обзор существующих методов стратификации. Во втором разделе формулируется проблема многокритериальной стратификации. В третьем разделе приводится оптимизационная задача формирования многокритериальной стратификации и новый алгоритм её решения. Четвертый раздел содержит описание реальных и сгенерированных данных, а также результаты экспериментов.

Работа велась при частичной финансовой поддержке международной Лаборатории анализа и выбора решений НИУ ВШЭ. Автор выражает благодарность д.т.н. Б.Г. Миркину за поставленную задачу.

### 1. Постановка проблемы стратификации

Множество из  $N$  объектов, оцененных по  $M$  критериям, необходимо разделить на  $K$  непересекающихся групп, упорядоченных между собой таким образом, чтобы объекты из одной группы были бы как можно более схожи по предпочтению с объектами из той же группы и в основном лучше, чем объекты из групп с большим номером. Такое упорядоченное разбиение будем называть стратификацией, а полученные группы объектов стратами. Оценки объектов по критериям будем записывать в виде критериальной матрицы  $X = \|x_{ij}\|$ , где  $i = 1, \dots, N$  – объекты или варианты,  $j = 1, \dots, M$  – критерии, а  $x_{ij}$  – значение  $j$ -го критерия для  $i$ -го объекта. Искомое множество страт обозначим  $S = \{S_1, \dots, S_K\}$ , где  $S_k$  – множество объектов, принадлежащих  $k$ -й страте ( $k = 1, \dots, K$ ). Объект из страты с номером  $k$  имеет более высокий ранг или предпочтительней чем объект из страты  $l$ , если  $k < l$ .

### 2. Критерий многомерной стратификации

В работе [6] предложена целевая функция для формирования оптимальной многокритериальной стратификации. Идеально, страты образуют параллельные гиперплоскости. Значения критериев  $i$ -го объекта удовлетворяют уравнению:

$$x_{i1}w_1 + x_{i2}w_2 + \dots + x_{iM}w_M = c_k + e_i,$$

если он принадлежит страте с номером  $k$ , где  $c_k \in \{c_1, c_2, \dots, c_K\}$  – искомые центры или «уровни» страт, искомые  $w_j$  – веса критериев,  $e_i$  – минимизируемая ошибка. Для получения стратификации, веса  $w$ , уровни страт  $c$  и разбиение  $S$  выбираются путем решения оптимизационной задачи (1):

$$\begin{cases} \sum_{k=1}^K \sum_{i \in S_k} \left( \sum_{j=1}^M x_{ij} w_j - c_k \right)^2 \rightarrow \min_{w, c, S} \\ \sum_{j=1}^M w_j = 1, w_j \geq 0 \end{cases} \quad (1)$$

Смысл задачи: найти такую комбинацию весовых коэффициентов  $w$ , и такие уровни  $c_k$  агрегированного критерия  $\sum_j w_j x_j$ , чтобы значения агрегированного критерия на всех объектах  $k$ -й страты были как можно ближе к  $c_k$  [6].

### 3. Алгоритм решения оптимизационной задачи многокритериальной линейной стратификации

Для формирования стратификации предлагаемый алгоритм решает задачу (1) используя подход чередующейся минимизации, который заключается в поиске оптимального решения по одной переменной при фиксированных других. На таком принципе основан, например, один из вариантов алгоритма  $k$ -средних [11]. Решение задачи (1) ищется чередованием трех последовательных шагов:

- 1) При фиксированных весах  $w_j$  и центрах  $c_k$  найти оптимальное разбиение;
- 2) При фиксированных весах  $w_j$  и разбиении  $S$  найти оптимальные центры;
- 3) При фиксированных центрах  $c_k$  и разбиении  $S$  найти оптимальные веса  $w_j$ .

На первых двух шагах решение получается следующим образом: оптимальное разбиение находится присвоением каждому объекту номера страты с ближайшим центром, а центры вычисляются как средние значения взвешенного критерия для объектов внутри страт. На третьем шаге оптимальные

веса критериев получаются из решения задачи квадратичного программирования с ограничениями. Поскольку на данном этапе разбиение фиксировано, каждому объекту приписывается соответствующий ему центр страты  $c_k(i) \in \{c_1, c_2, \dots, c_K\}$ ,  $i = 1 \dots N$ .

Для удобства перепишем (1) в более компактном виде. Воспользуемся соотношением  $c_k(i) = \sum_{j=1}^M x_{ij} w_j$ , поскольку веса в сумме равны единице. Затем перепишем выражение внутри скобок (1) в следующем виде:

$$\sum_{k=1}^K \sum_{i \in S_k} \left( \sum_{j=1}^M x_{ij} w_j - c_k(i) \right)^2 = \sum_{k=1}^K \sum_{i \in S_k} \left( \sum_{j=1}^M x_{ij} w_j - c_k(i) w_j \right)^2 = \sum_{k=1}^K \sum_{i \in S_k} \left( \sum_{j=1}^M (x_{ij} - c_k(i)) w_j \right)^2 = \sum_{k=1}^K \sum_{i \in S_k} \left( \sum_{j=1}^M w_j \bar{x}_{ij} \right)^2$$

Сделав замену переменной  $\bar{x}_{ij} = (x_{ij} - c_k(i))$  и введя обозначение  $\bar{X} = \|\bar{x}_{ij}\|$  перепишем (1) в матричном виде:

$$\begin{cases} w^T \bar{X}^T \bar{X} w \rightarrow \min \\ \sum_{j=1}^M w_j = 1 \\ w_j \geq 0 \end{cases} \quad (2)$$

Для решения (2) можно воспользоваться одним из известных алгоритмов квадратичного программирования, например, методом активного множества (active-set algorithm) [7], имплементированным в пакете Matlab 7. Предлагаемый алгоритм стратификации на основе квадратичного программирования Linstrat-Q приведен ниже.

#### Алгоритм Linstrat-Q

На входе:

- Объекты  $x_i$ ,  $i = 1 \dots N$ ;
- Число страт  $K$ ;
- Число итераций  $T$ .

На выходе:

- Веса  $w$ ;
- Центры страт  $c$ ;
- Разбиение  $S$ .

1. Инициализировать веса  $w$  и центры страт  $c$ . Сгенерировать веса  $w$  случайно такие, что  $w_j \geq 0$ ,  $j = 1 \dots M$  и  $\sum_{j=1}^M w_j = 1$ . Вычислить свертку критериев с весами. Центры страт сгенерировать случайно равномерно из диапазона от минимального значения свертки критериев до максимального.

2. По заданным весам и центрам найти оптимальное разбиение:

$$x_i \in S_k, \text{ где } k = \operatorname{argmin}_k \left( \sum_{j=1}^M x_{ij} w_j - c_k \right)^2,$$

$$k = 1 \dots K, i = 1 \dots N.$$

3. По заданным весам и разбиению на страты найти оптимальные центры:

$$c_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} \sum_{j=1}^M x_{ij} w_j.$$

4. По заданным центрам и разбиению найти оптимальные веса, решив задачу квадратичного программирования (2).

5. Завершить, если сделано  $T$  итераций, иначе перейти к п. 2.

Вопрос об инициализации алгоритмов, строящих разбиения, является одним из ключевых в кластерном анализе. В зависимости от начальных значений весов  $w$  и центров  $c$  могут получаться различные разбиения. Данная работа следует традиционному эвристическому подходу – задать начальные параметры случайно, найти решение для каждой такой инициализации, а затем выбрать решение, для которого значение целевой функции будет минимально. Другой так же важный вопрос – об определении оптимального числа страт – в данной работе не рассматривается

Продемонстрируем работу алгоритма по шагам на простом примере. Рассмотрим шесть объектов оцененных по двум критериям  $x_1 = (0,1)$ ,  $x_2 = (1,0)$ ,  $x_3 = (1,2)$ ,  $x_4 = (2,1)$ ,  $x_5 = (2,3)$ ,  $x_6 = (3,2)$ .

Шаг 1. Произведем инициализацию весов и центров. Допустим, сгенерированные значения равны  $w = (0.2, 0.8)$ ,  $c_3 = 0.3$ ,  $c_2 = 1.9$ ,  $c_1 = 2.4$ .

Шаг 2. Вычисляем свертку критериев с весами. Получаем значения для каждого объекта соответственно 0.8, 0.2, 1.8, 1.2, 2.8, 2.2. Далее для первого объекта вычисляем квадрат разности значения свертки и значения центра каждой страты  $(c_k - x_1 w^T)^2$ ,  $k = 1, 2, 3$ . Минимальным является квадрат разности для третьей страты  $(c_k - x_1 w^T)^2 = (0.3 - 0.8)^2 = 0.25$ , следовательно, назначаем первый объект на страту с номером 3. Аналогично производим назначения для остальных объектов. Получаем, что объекты принадлежат соответственно стратам с номерами 3, 3, 2, 2, 1, 1.

Шаг 3. Теперь для каждой из страт вычисляем центры как средние значения взвешенных критериев объектов, принадлежащих заданной страте. Таким образом,

$$c_3 = \frac{1}{2} \cdot (0.8 + 0.2) = 0.5, c_2 = \frac{1}{2} \cdot (1.8 + 1.2) = 1.5,$$

$$c_1 = \frac{1}{2} \cdot (2.8 + 2.2) = 2.5.$$

Шаг 4. Вычисляем матрицу  $\bar{X}$  для построения задачи квадратичного программирования (2):

$$\bar{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 2 \\ 2 & 1 \\ 2 & 3 \\ 3 & 2 \end{pmatrix} - \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \\ 1.5 & 1.5 \\ 1.5 & 1.5 \\ 2.5 & 2.5 \\ 2.5 & 2.5 \end{pmatrix} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \\ -0.5 & 0.5 \\ 0.5 & -0.5 \\ -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$$

Матрица для задачи квадратичного программирования (2) будет иметь вид:

$$\bar{X}^T \bar{X} = \begin{pmatrix} 1.5 & -1.5 \\ -1.5 & 1.5 \end{pmatrix}$$

Решение задачи квадратичного программирования дает веса  $w = (0.5, 0.5)$ . Повторяем три вышеописанных шага заданное число итераций.

На шаге 2 для каждого из  $N$  объектов производится свертка с весами по  $M$  критериям и вычисляется наименьшая из  $K$  квадратов разностей, значит оценка сложности этого шага в худшем случае  $O(NMK)$ . На шаге 3 вычисляются средние значения взвешенных критериев внутри страт. Сложность этого шага  $O(N)$ . Далее, на шаге 4 требуется  $O(NM)$  операций для вычисления матрицы  $\bar{X}$ . Произведение  $\bar{X}^T \bar{X}$  требует  $O(NM^2)$  операции. Обозначим  $O(f(M))$  оценку сложности решения задачи квадратичного программирования (2), где  $f$  зависит от выбранного алгоритма оптимизации. При заданном  $T$  числе итераций оценка сложности алгоритма будет  $O((NM(M+K)+f(M))T)$ .

#### 4. Экспериментальное сравнение алгоритмов стратификации

Для апробации рассматриваемых в работе алгоритмов и их экспериментального сравнения были взяты искусственные и реальные данные. Предложен усовершенствованный алгоритм генерации стратифицированных данных, позволяющий всесторонне контролировать конфигурацию страт. В качестве реальных данных были взяты библиометрические показатели журналов и стран.

Ниже, в табл. 1 перечислены методы стратификации, используемые в работе для экспериментального сравнения. Более подробный обзор рассматриваемых методов приведен в работе [6].

Таблица 1.

Список методов стратификации и их аббревиатур, используемых в данной работе

Метод стратификации	Аббревиатура	Источник
Метод линейной стратификации Linstrat-Q с использованием квадратичного программирования	LSQ	Эта работа
Метод линейной стратификации Linstrat на основе эволюционной минимизации	LS	[6]
Стратификация с помощью правила Борда (Borda count)	BC	[2]
Метод ABC- классификации на основе линейной оптимизации весов (Linear weights optimization)	LWO	[13]
Ранжирование по влиянию (Authority ranking)	AR	[19]
Стратификация объединением границ Парето (Paretostrat)	PS	[6]

Все методы были имплементированы в среде Matlab 7. Для решения задачи квадратичного программирования (2) использовалась функция quadprog из библиотеки Matlab Optimization Toolbox. Для решения задачи линейного программирования для метода LWO использовалась функция linprog из вышеупомянутой библиотеки. Для BC, LWO, AR после получения одномерного ранжирования стратификация производилась по численным значениям агрегированного критерия. Для этого использовался алгоритм  $k$ -средних. Случайная инициализация для LS, LSQ, а также  $k$ -средних для BC, LWO, AR выполнялась 100 раз, и записывался результат, дающий наименьшее значение целевой функции. Число итераций  $T$  для алгоритма LS и LSQ в экспериментах было установлено равным 100.

Для оценки качества стратификации при контролируемом экспериментальном исследовании на искусственных данных использовался показатель точности стратификации, т. е. отношение объектов с правильно определенным номером страты к общему числу объектов:

$$accuracy = \frac{N_{correct}}{N} \quad (3)$$

На реальных данных такая мера качества не применима, поскольку заведомо не известно к какой страте относится объект. Мы полагаем, что хорошее многокритериальное разбиение должно быть согласованно с разбиением по каждому отдельно взятому критерию. Поэтому для оценки качества работы алгоритма на реальных данных мы ис-

пользовали среднее расстояние от многокритериальной стратификации до 3 однокритериальных стратификаций по каждому отдельно взятому критерию. В качестве расстояния между стратификациями  $S$  и  $R$  использовалось нормированное расстояние Кемени-Снелла [4, 10]. Для стратификации  $S$ , заданной на объектах  $x_1, \dots, x_N$  строилась матрица:

$$s_{ij} = \begin{cases} 1, S(x_i) > S(x_j) \\ 0, S(x_i) = S(x_j) \\ -1, S(x_i) < S(x_j) \end{cases} \quad (4)$$

где  $S(x)$  обозначает номер страты, которой принадлежит объект  $x$ . Аналогично формируется матрица стратификации  $R$ . Расстояние вычисляется по формуле:

$$d_{RS} = \frac{1}{2N(N-1)} \sum_{i,j=1}^N |R_{ij} - S_{ij}|. \quad (5)$$

#### 4.1 Модель генерации стратифицированных данных

Для того чтобы обеспечить возможность проведения контролируемых экспериментов по сравнению алгоритмов, в работе [6] был предложен случайный механизм генерации стратифицированных данных.

В данной работе предлагается усовершенствованная версия алгоритма генерации, позволяющая более разносторонне контролировать геометрическую конфигурацию страт. Алгоритм генерирует страты в виде параллельных гиперплоскостей. Более точно, геометрия страт задается различными параметрами, к которым относятся: весовые коэффициенты  $w$ , уровни страт  $c$ , интенсивности страт  $\theta$ , размах страт  $\varphi$  и толщина страт  $\sigma$ . Смысл этих величин раскрывается в нижеследующем алгоритме генерации стратифицированных данных. Список значений, задаваемых по умолчанию параметров генерации, приведен в табл. 1 Приложения 1. На рис. 1 наглядно продемонстрировано то, как выглядят страты в зависимости от параметров генерации. Алгоритм генерации описан ниже.

#### Алгоритм генерации стратифицированных данных

На входе:

- Число объектов  $N$ , размерность  $M$  и число страт  $K$ ;
- Уровни страт  $c$ ;
- Весовые коэффициенты (ориентация страт)  $w$ ;

- Толщина страт  $\sigma$ ;
- Интенсивности страт  $\theta$ ;
- Размах страт  $\varphi$ .

На выходе:

- Значения критериев для каждого объекта;
- Индекс страты для каждого объекта.

1. Выбрать номер страты из мультиномиального распределения  $k \sim M(\theta_1, \theta_2, \dots, \theta_K)$
2. Выбрать величину, моделирующую значение агрегированного критерия (значения свертки критериев с весами) на объекте, из нормального распределения  $r \sim N(c_k, \sigma)$
3. Сгенерировать  $M-1$  координат из равномерного распределения:  $x_j \sim U(c_k(1-\varphi), c_k(1+\varphi)/w_j), j = 1 \dots M-1$ .
4. Вычислить последнюю  $M$ -ю координату из уравнения гиперплоскости:  $x_M = (r - w_1x_1 + w_2x_2 + \dots + w_{M-1}x_{M-1})/w_M$ .
5. Завершить, если сгенерировано  $N$  объектов, иначе перейти к п. 1.

#### 4.2. Реальные данные.

##### Библиометрические показатели журналов и стран

Реальные данные были взяты с портала scimagojr.com, разработанного исследовательской группой SCImago (<http://www.scimago.com/>). Этот портал содержит в открытом доступе библиометрические показатели журналов и стран, полученные на основе данных из базы Scopus (<http://www.elsevier.com/online-tools/scopus>). Данные из этого источника широко используются в научных разработках (см., например, [1, 17, 18]).

##### 4.2.1 Данные о библиометрических показателях журналов

Рассмотрим библиометрические показатели научных журналов из раздела Artificial Intelligence за 2012 год. Всего в этом разделе 118 журналов. В качестве критериев оценки престижа журнала рассмотрим три наиболее популярных библиометрических показателя:

- 1) Индекс SJR (Scientific Journal Ranking) [8]. Его значение отражает среднее число посещений журнала неким условным читателем, со-

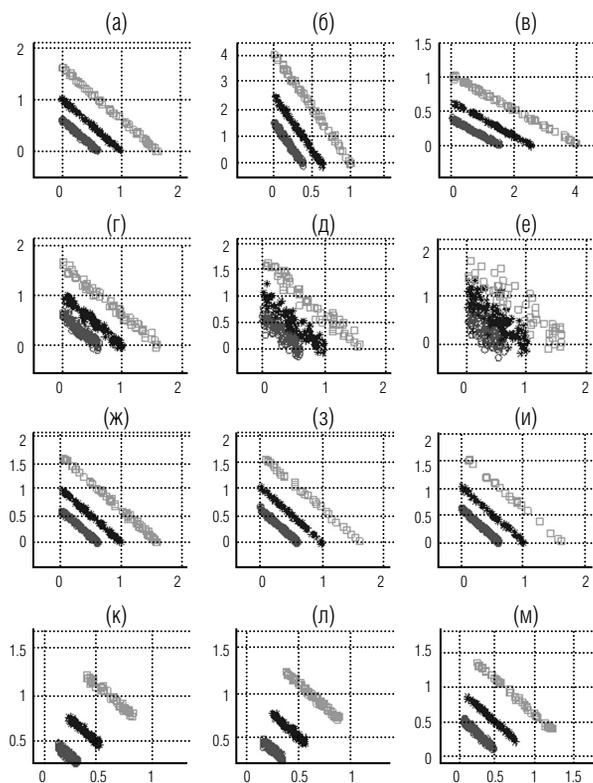


Рис. 1. Линейные страты для различных комбинаций параметров генерации в двумерном случае,  $M = 2$ .

Приведены следующие варианты:

- (1) страты в зависимости от ориентации  $w$   
(а)  $w = (0.5, 0.5)$ , (б)  $w = (0.8, 0.2)$ , (в)  $w = (0.2, 0.8)$ ;
- (2) страты в зависимости от толщины  
(г)  $\sigma = 0.05$ , (д)  $\sigma = 0.1$  (е)  $\sigma = 0.2$ ;
- (3) страты в зависимости от интенсивности  
(ж)  $\theta = (0.5, 0.3, 0.2)$ , (з)  $\theta = (0.7, 0.2, 0.1)$ , (и)  $\theta = (0.8, 0.15, 0.05)$ ;
- (4) страты в зависимости от размаха  
(к)  $\Phi = 0.05$ , (л)  $\Phi = 0.1$ , (м)  $\Phi = 0.5$ .

вершающим случайные переходы по ссылкам на документы журналов. Данный индекс развивает идею, лежащую в основе известного алгоритма пейджранк [12].

2) Индекс Хирша (H) [9]. Этот индекс характеризует количество статей  $h$ , опубликованных журналом, таких, что число их цитирований выше значения  $h$ .

3) Импакт-фактор журнала (I) [20]. Этот индекс рассчитывается отнесением числа цитирований статей, опубликованных в журнале в таком-то году, за последующие два года, к числу статей, опубликованных в журнале в том самом году.

Более подробное обсуждение вышеперечисленных и других наукометрических показателей, их достоинств и недостатков может быть найдено в ряде работ, например, в [1, 3, 5].

#### 4.2.2 Данные о библиометрических показателях публикационной активности стран

Из упомянутого выше источника были взяты данные о публикационной активности 102 стран за 2012 год, в предметной области – искусственный интеллект (Artificial intelligence). Критерии, используемые для оценки стран, следующие:

- 1) Общее число документов опубликованных за 2012 (D);
- 2) Число цитируемых документов, опубликованных в 2012 году (CD);
- 3) Общее количество цитирований в 2012 году, полученных документами, опубликованными в этом же году (C).
- 4) Самоцитирование документов в 2012 году (country self-citations) (SC);
- 5) Среднее число цитирований в 2012 году документов, опубликованных в этом году (CPD);
- 6) H-индекс на уровне страны (H).

#### 4.2.3 Нормировки данных

Для реальных данных мы использовали два способа нормировки: стандартная нормировка (6) (приведение значений критериев к диапазону [0,1]), и статистическая нормировка (7) (приведение значений к нулевому среднему и единичному стандартному отклонению):

$$z_{ij} = \frac{x_{ij} - \min_j(x_j)}{\max_j(x_j) - \min_j(x_j)} \quad (6)$$

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (7)$$

В формуле (6)  $\max_j(x_j)$  и  $\min_j(x_j)$  максимальное и минимальное значение по  $j$ -му критерию. В формуле (7)  $\mu_j$  и  $\sigma_j$  среднее значение и стандартное отклонение значений  $j$ -го критерия.

#### 4.3 Результаты экспериментов на искусственных данных

В этом разделе приведены результаты экспериментов по стратификации, в зависимости от параметров генерации страт: размерности, общего количества сгенерированных объектов, интенсивности, размаха и толщины страт. Критерий оценки качества работы алгоритма – точность стратификации (3). Результаты экспериментов приведены в табл. 2-5 Приложения 1.

Предлагаемый алгоритм LSQ оказался достаточно устойчивым к увеличению размерности (табл. 2, Приложение 1). В тоже время точность стратификации методом LS значительно снизилась уже при размерности больших пяти. Точность стратификации заметно снизилась с ростом размерности также для методов LWO и PS. Алгоритм AR показал обратную тенденцию: точность стратификации с увеличением размерности возросла.

Снижение точности стратификации для LSQ при увеличении размерности может быть компенсировано увеличением объема выборки (табл. 3, Приложение 1). Такой же эффект «компенсации» наблюдается и для LWO. Методы AR и BC являются относительно нечувствительными к объему данных. А для LS и PS снижение точности стратификации при увеличении размерности не удается компенсировать увеличением объема данных.

Алгоритмы линейной стратификации LSQ и LS показали высокую устойчивость к изменению интенсивностей страт  $\theta$  по сравнению с другими алгоритмами (табл. 4, Приложение 1).

При малых значениях размаха  $\varphi$  все алгоритмы показали достаточно высокие результаты (табл. 5, Приложение 1). Особенно сильно этот параметр влияет на работу алгоритма PS. Не так заметно, но, тем не менее, снижается и точность для AR и BC. В меньшей степени размах влияет на LWO и совсем не влияет на качество работы LS и LSQ.

Утолщение страт влияет на качество работы всех алгоритмов (табл. 6, Приложение 1). Алгоритмы LSQ, LS, показывающие отличную точность при небольших значениях  $\sigma$ , с увеличением  $\sigma$  начинают уступать алгоритму LWO, который оказался наиболее устойчивым к увеличению толщины страт.

Таким образом, практически во всех экспериментах LSQ входит в группу лидеров. Это выделяет его из всех других алгоритмов, которые могут оказаться чувствительными к тому или иному параметру.

#### 4.4. Результаты экспериментов на реальных данных

В этом разделе приведены результаты экспериментов по стратификации на реальных данных, описанных в п. 4.2. Рассмотрена стратификация 118 научных журналов и 102 стран по библиометрическим показателям. Обсуждается согласованность разбиений. Оценкой степени согласованности служит среднее расстояние Кемени-Снелла

(5) между многокритериальной стратификацией и стратификациями по каждому отдельному критерию.

##### 4.4.1 Результаты стратификации академических журналов по библиометрическим показателям

Как видно из табл. 2, Приложения 2, наиболее согласованное разбиение для обоих типов нормировки получилось методами линейной стратификации LS и LSQ. Метод AR нашел согласованное разбиение только при стандартной нормировке.

Наш алгоритм LSQ сформировал одинаковые стратификации для обеих нормировок. При этом наибольшие веса получили индекс SJR и импакт-фактор (см. табл. 1, Приложения 2). То есть по этим двум критериям можно получить хорошо стратифицированное множество журналов. В то же время индекс Хирша получил небольшой вес – таким образом, алгоритм как бы присоединяется к критике этого индекса (см., например, сборник «Игра в цифри» 2011). В первую страту вошли 6 журналов, во вторую 42 и в третью 70. В первой страте журналы, высоко ценимые в сообществе исследователей:

1. IEEE Transactions on Pattern Analysis and Machine Intelligence;
2. International Journal of Computer Vision;
3. Foundations and Trends in Machine Learning;
4. ACM Transactions on Intelligent Systems and Technology;
5. IEEE Transactions on Evolutionary Computation;
6. IEEE Transactions on Fuzzy Systems.

Несколько неожиданным является попадание в эту группу относительно нового журнала Foundations and Trends in Machine Learning. Этот журнал публикует высококачественные монографические обзоры по актуальным проблемам.

##### 4.4.2 Результаты стратификации стран по библиометрическим показателям

Так же, как и в предыдущем эксперименте, наиболее согласованные разбиения получились по методам LS и LSQ (см. табл. 2, Приложения 3). Веса критериев записаны в табл. 1, Приложения 3, из них ненулевые веса получили, главным образом, два критерия: самоцитирование и индекс Хирша. Остальные критерии на данной выборке ведут себя неустойчиво, и не позволяют разделить ее на

«параллельные» слои. Это, прежде всего, характеристики общего количества статей (критерии 1 и 2 из п. 4.2.2), и характеристики цитирования в том же году, также носящие случайный характер: действительно, для действенного учета новой публикации необходимо время на развитие и адаптацию ее идей, а также на публикацию полученных результатов. Расчеты для данных за 2009 год (при этом в расчет берется цитирование работ за более долгий период 2009 - 2012) приводят к аналогичным результатам, при этом среднее число цитирований получает несколько больший вес – 13% вместо 5%, а веса критериев 1 и 2 остаются нулевыми.

Первую страту составили две страны: Китай, США. Во второй страте оказались 17 стран: Испания, Англия, Франция, Тайвань, Япония, Индия, Германия, Канада, Италия, Южная Корея, Австралия, Гонконг, Голландия, Сингапур, Швейцария, Израиль. Остальные 83 страны сформировали третью страту.

#### Заключение

В работе рассмотрена проблема многокритериальной стратификации на основе автоматического определения весовых коэффициентов критериев таким образом, чтобы получаемые страты образовывали компактные множества на оси обобщенно-

го критерия. Предложен алгоритм, использующий квадратичную оптимизацию. Этот алгоритм экспериментально сравнен с группой существующих подходов. Как оказалось, в большинстве случаев предложенный алгоритм приводит к наилучшим решениям. Применение алгоритма к реальным данным приводит к достаточно хорошо интерпретируемым результатам.

В дальнейшем мы планируем проверить работу алгоритма на различных реальных данных, особенно в применении к динамическим рядам многомерных критериев с целью дальнейшего изучения проблем интерпретации и устойчивости получаемых решений. Заслуживает внимания, в частности, проблема устойчивости алгоритма относительно изменения толщины страт.

Существенным недостатком подхода к стратификации на основе рассматриваемого в работе критерия линейной стратификации является то, что данный критерий не учитывает расстояния между стратами. То есть, из нескольких решений, позволяющих получить одинаково плотные группировки на оси взвешенного критерия, не будет сделан выбор в пользу того, которое обеспечивает, максимальный зазор между стратами. В настоящее время исследуется возможность сформулировать критерий стратификации таким образом, чтобы преодолеть вышеупомянутый недостаток. ■

#### Литература

1. Алескерев Ф.Т., Писляков В.В., Субочев А.Н. Построение рейтингов журналов по экономике с помощью методов теории коллективного выбора. WP7/2013/03. – М.: Изд. дом ВШЭ, 2013.
2. Алескерев Ф.Т., Хабина Э.Л., Шварц Д.А. Бинарные отношения, графы и коллективные решения. М.: ГУ-ВШЭ, 2006.
3. Игра в цыфирь, или как теперь оценивают труд ученого. М.: Московский Центр непрерывного математического образования, 2011.
4. Миркин Б.Г. Проблема группового выбора. М.: Наука, 1974.
5. Миркин Б.Г. О понятии научного вклада ученого и его измерителях // Управление большими системами: сборник трудов. 2013. С. 1-16.
6. Миркин Б.Г., Орлов М.А. Методы многокритериальной стратификации и их экспериментальное сравнение. WP7/2013/03. М.: Изд. дом ВШЭ, 2013.
7. Gill P.E., Murray W., Wright M.H. Numerical linear algebra and optimization. Vol. 1. Addison Wesley, 1991.
8. Gonzalez-Pereira B., Guerrero-Bote V., Moya-Anegon F. A new approach to the metric of journals scientific prestige: The SJR indicator // Journal of Informetrics. 2010. Vol. 4, issue 3. P. 379-391.
9. Hirsch J.E. An index to quantify an individual's scientific research output. 2005 (Retrieved from arXiv 05 February 2014).
10. Kemeny J., Snell L. Mathematical models in the social sciences. Boston: Ginn, 1962.
11. Mirkin B.G. Clustering: A data recovery approach. CRC Press, 2012.
12. Page L., Brin S., Motwani R., Winograd T. The PageRank citation ranking: Bringing order to the Web. Technical Report. Stanford InfoLab, 1999.

13. Ramanathan R. Inventory classification with multiple criteria using weighted linear optimization // Computers and Operations Research. 2006. No. 33. P. 695-700.
14. SCImago Journal & Country Ranking. (2007). SJR – SCImago Journal & Country Rank. <http://www.scimagojr.com>. (дата обращения 14.01.2014).
15. SCImago Lab. <http://www.scimagolab.com/> (дата обращения 14.01.2014).
16. Scopus. <http://www.elsevier.com/online-tools/scopus> (дата обращения 22.01.2014).
17. Siebelt M., Siebelt T., Pilot P., Bloem R.M., Bhandari M., Poolman R.W. Citation analysis of orthopedic literature; 18 major orthopedic journals compared for Impact Factor and SCImago // BMC Musculoskeletal Disorders. 2010.
18. Spreckelsen C., Deserno T.M., Spitzer K. Visibility of medical informatics regarding bibliometric indices and databases // BMC Medical Informatics and Decision Making. 2011.
19. Sun Y., Han J., Zhao P., Yin Z., Cheng H., Wu T. RankClus: Integrating clustering with ranking for heterogeneous information network analysis // Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (EDBT 2009). March 23-26 2009, Saint-Petersburg, Russia. P. 565-576.
20. Garfield E. The Thomson Reuters Impact Factor. Thomson Reuters, 1994.

**ПРИЛОЖЕНИЕ 1.**

**Результаты экспериментального исследования алгоритмов стратификации на искусственных данных**

Таблица 1.

Значения параметров генерации страт по умолчанию

Параметр генерации страт	Значение по умолчанию
Число объектов, $N$	$N = 300$
Число критериев, $M$	$M = 5$
Число страт, $K$	$K = 3$
Центры страт, $c$	$c = (0.3, 0.5, 0.8)$
Интенсивности страт, $\theta$	$\theta = (0.33, 0.33, 0.33)$
Веса критериев, $w$	$w = (0.2, 0.2, 0.2, 0.2, 0.2)$
Толщина страт, $\sigma$	$\sigma = 0.01$
Размах страт, $\varphi$	$\varphi = 1$

Таблица 2.

Точность стратификации (среднее значение и стандартное отклонение для 10 генераций) различными методами в зависимости размерности данных  $M$

Метод	Размерность данных		
	$M = 3$	$M = 5$	$M = 20$
LSQ	<b>1.00±0.00</b>	<b>1.00±0.00</b>	0.84±0.02
LS	<b>1.00±0.00</b>	<b>1.00±0.00</b>	0.44±0.08
BC	0.90±0.02	0.73±0.03	<b>0.88±0.01</b>
LWO	<b>1.00±0.00</b>	0.98±0.01	0.45±0.04
AR	0.60±0.02	0.61±0.04	<b>0.88±0.03</b>
PS	<b>0.99±0.00</b>	0.36±0.02	0.32±0.04

Таблица 3.

Точность стратификации (среднее значение и стандартное отклонение для 10 генераций) в зависимости от количества сгенерированных объектов  $N$  при размерности  $M = 20$

Метод	Число сгенерированных объектов		
	$N = 200$	$N = 300$	$N = 500$
LSQ	0.65±0.27	0.75±0.22	<b>1.00±0.00</b>
LS	0.49±0.11	0.43±0.11	0.44±0.10
BC	<b>0.87±0.03</b>	<b>0.88±0.02</b>	0.88±0.02
LWO	0.38±0.04	0.44±0.03	0.63±0.04
AR	<b>0.88±0.05</b>	<b>0.89±0.04</b>	0.88±0.02
PS	0.33±0.04	0.33±0.02	0.33±0.02

Таблица 4.

Точность стратификации (среднее значение и стандартное отклонение для 10 генераций) в зависимости от интенсивностей  $\theta$

Метод	Паттерн интенсивности страт		
	$\theta = (0.33, 0.33, 0.33)$	$\theta = (0.5, 0.3, 0.2)$	$\theta = (0.6, 0.4, 0.1)$
LSQ	<b>1.00±0.00</b>	<b>1.00±0.00</b>	<b>1.00±0.00</b>
LS	<b>1.00±0.00</b>	<b>1.00±0.00</b>	<b>1.00±0.00</b>
BC	0.75±0.03	0.77±0.03	0.56±0.11
LWO	0.99±0.01	0.95±0.03	0.88±0.05
AR	0.60±0.04	0.67±0.03	0.65±0.06
PS	0.36±0.02	0.20±0.03	0.09±0.02

Таблица 5.  
Точность стратификации  
(среднее значение и стандартное отклонение  
для 10 генераций данных) в зависимости  
от размаха страт  $\varphi$

Метод	Размах		
	$\varphi = 0.01$	$\varphi = 0.1$	$\varphi = 1$
LSQ	<b>1.00±0.00</b>	<b>1.00±0.00</b>	<b>1.00±0.00</b>
LS	<b>1.00±0.00</b>	<b>1.00±0.00</b>	<b>1.00±0.00</b>
BC	0.97±0.01	0.95±0.01	0.74±0.03
LWO	0.99±0.00	0.99±0.01	0.98±0.01
AR	0.67±0.04	0.65±0.05	0.59±0.02
PS	0.96±0.02	0.86±0.04	0.36±0.02

Таблица 6.  
Точность стратификации  
(среднее значение и стандартное отклонение  
для 10 генераций данных) в зависимости  
от толщины  $\sigma$

Метод	Толщина страт		
	$\sigma = 0.01$	$\sigma = 0.1$	$\sigma = 0.2$
LSQ	<b>1.00±0.00</b>	<b>0.99±0.01</b>	0.76±0.09
LS	<b>1.00±0.00</b>	<b>0.99±0.01</b>	0.77±0.09
BC	0.76±0.03	0.73±0.03	0.71±0.03
LWO	<b>0.98±0.01</b>	0.96±0.01	<b>0.90±0.02</b>
AR	0.59±0.03	0.58±0.04	0.59±0.03
PS	0.35±0.02	0.38±0.02	0.42±0.03

## ПРИЛОЖЕНИЕ 2.

### Анализ многокритериальной стратификации журналов по библиометрическим показателям

Таблица 1.  
Веса критериев важности  
научных журналов, найденные методом  
линейной стратификации LSQ

Критерий	Веса критериев, найденные при стратификации	
	Стандартная нормировка (7)	Статистическая нормировка (8)
Индекс SJR	0.38	0.47
Индекс Хирша, H	0.14	0.14
Импакт фактор, I	0.47	0.39

Таблица 2.  
Средние значения расстояний  
Кемени-Снелла между упорядоченными  
разбиениями журналов по отдельным критериям  
и разбиениями, найденными методами  
многокритериальной стратификации

Метод	LSQ	LS	BC	LWO	AR	PS
Среднее расстояние. (Стандартная нормировка)	<b>0.12</b>	<b>0.12</b>	0.17	0.13	<b>0.12</b>	0.23
Среднее расстояние. (Статистическая нормировка)	<b>0.12</b>	<b>0.12</b>	0.17	0.15	0.20	0.23

## ПРИЛОЖЕНИЕ 3.

### Анализ многокритериальной стратификации стран по публикационной активности

Таблица 1.  
Веса критериев оценки публикационной  
активности стран, найденные методом  
линейной стратификации LSQ

Критерий	Веса критериев, найденные при стратификации	
	Стандартная нормировка (7)	Статистическая нормировка (8)
1	0	0
2	0	0
3	0	0
4	<b>0.63</b>	<b>0.52</b>
5	0.05	0.07
6	0.33	0.41

Таблица 2.  
Средние значения расстояний  
Кемени-Снелла между упорядоченными  
разбиениями стран по отдельным критериям  
и разбиениями, найденными методами  
многокритериальной стратификации

Метод	LSQ	LS	BC	LWO	AR	PS
Среднее расстояние. (Стандартная нормировка)	<b>0.10</b>	<b>0.10</b>	0.26	0.21	0.16	0.18
Среднее расстояние. (Статистическая нормировка)	<b>0.10</b>	<b>0.10</b>	0.26	0.17	0.12	0.18

# AN ALGORITHM FOR MULTICRITERIA STRATIFICATION

**Mikhail ORLOV**

Post-graduate Student, Department of Data Analysis and Artificial Intelligence,  
Faculty of Computer Science,

National Research University Higher School of Economics

Address: 20, Myasnitskaya street, Moscow, 101000, Russian Federation

E-mail: [ormian@mail.ru](mailto:ormian@mail.ru)

*This paper elaborates an approach to the problem of multicriteria ranking referred to as multicriteria stratification. The target of stratification is an ordered partition with predefined number of classes – strata rather than a complete ranking of the set of objects. Ranking is computed by means of linear convolution of criteria with some weights. These weights are based on assumption that data can fit some linear structure so that «parallel» layers can be identified – strata.*

*In the paper [6] the authors formulated the problem of multicriteria stratification as a task of minimization of a cost function depending on criteria weights; however the algorithm proposed in that paper to address the emerging task based on random searching has demonstrated low performance in comparison to some other stratification approaches.*

*In this paper a new algorithm based on quadratic programming is proposed to optimize the multicriteria stratification target function. A more sophisticated synthetic data generator for a comparative study of the stratification algorithm has been developed. The new data generator has more parameters to tune and allows more flexible control of geometry of synthetic strata: orientation, thickness, spread and intensity of layers that enables to pay due regard to real data structure.*

*The novel algorithm has been compared experimentally with existing stratification approaches by involving synthetic data, and its competitiveness has been shown in the majority of case studies. Two real-world datasets have been processed – bibliometrical indicators of 118 scientific journals and parameters of publication activities of 102 countries. The new algorithm applied to handle these data has produced sensible and well interpretable outputs. Furthermore, on these data the proposed algorithm found the most coherent multicriteria stratification to those computed by each single criterion.*

**Key words:** stratification, rank aggregation, multicriteria, weighted sum, quadratic programming, optimization, bibliometrics.

## References

1. Aleskerov F.T., Pisljakov V.V., Subochev A.N. (2013) *Postroenie rejtingov zhurnalov po jekonomike s pomoshh'ju metodov teorii kolektivnogo vybora* [Rankings of economic journals constructed by the Social Choice Theory methods]. WP7/2013/03. – Moscow: HSE. (in Russian)
2. Aleskerov F.T., Habina E.L., Shvarc D.A. (2006) *Binarnye otoshenija, grafy i kolektivnye reshenija* [Binary relations, graphs and group choice]. Moscow: HSE. (in Russian)
3. MCNMO (2011) *Igra v cyfir', ili kak teper' ocenivajut trud uchjonogo* [The game of numbers, or how scientific research is evaluated today] // Moscow: MCNMO. (in Russian)
4. Mirkin B.G. (1974) *Problema gruppovogo vybora* [Group choice problem]. Moscow: Nauka. (in Russian)
5. Mirkin B.G. (2013) O ponjatii nauchnogo vklada i ego izmeriteljah [On the concept of scientific contribution and its metrics] // *Upravlenie bol'shimi sistemami: sbornik trudov* [Big systems management: Proceedings], pp. 1-16. (in Russian)
6. Mirkin B.G., Orlov M.A. (2013) *Metody mnogokriterial'noj stratifikacii i ih jeksperimental'noe sravnenie* [Methods for multicriteria stratification and experimental comparisons]. WP7/2013/03. – Moscow: HSE. (in Russian)
7. Gill P.E., Murray W., Wright M.H. (1991) *Numerical linear algebra and optimization*, vol. 1. Addison Wesley.
8. Gonzalez-Pereira B., Guerrero-Bote V., Moya-Anegon F. (2010) A new approach to the metric of journals scientific prestige: The SJR indicator // *Journal of Informetrics*, vol. 4, no. 3, pp. 379-391.
9. Hirsch J.E. (2005) *An index to quantify an individual's scientific research output* (Retrieved from arXiv 05 February 2014).
10. Kemeny J., Snell L. (1962) *Mathematical models in the social sciences*. Boston: Ginn.

11. Mirkin B.G. (2012) *Clustering: A data recovery approach*. CRC Press.
12. Page L., Brin S., Motwani R., Winograd T. (1999) *The PageRank citation ranking: Bringing order to the Web*. Technical Report. Stanford InfoLab.
13. Ramanathan R. (2006) Inventory classification with multiple criteria using weighted linear optimization // *Computers and Operations Research*, no. 33, pp. 695-700.
14. SCImago Journal & Country Ranking. (2007). SJR — SCImago Journal & Country Rank. <http://www.scimagojr.com>. (Accessed 14 January 2014)
15. SCImago Lab. <http://www.scimago.com/> (Accessed 14 January 2014).
16. Scopus. <http://www.elsevier.com/online-tools/scopus> (Accessed 22 January 2014).
17. Siebelt M., Siebelt T., Pilot P., Bloem R.M., Bhandari M., Poolman R.W. (2010) *Citation analysis of orthopedic literature; 18 major orthopedic journals compared for Impact Factor and SCImago* // *BMC Musculoskeletal Disorders*.
18. Spreckelsen C., Deserno T.M., Spitzer K. (2011) *Visibility of medical informatics regarding bibliometric indices and databases* // *BMC Medical Informatics and Decision Making*.
19. Sun Y., Han J., Zhao P., Yin Z., Cheng H., Wu T. (2009) RankClus: integrating clustering with ranking for heterogeneous information network analysis // *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (EDBT 2009)*. March 23-26 2009, Saint-Petersburg, Russia, pp. 565-576.
20. Garfield E. (1994) *The Thomson Reuters Impact Factor*. Thomson Reuters.

# РАСЧЕТНАЯ МОДЕЛЬ ИЗМЕРЕНИЯ ЭЛЕКТРОЭНЕРГИИ В ИНФОРМАЦИОННЫХ БИЛЛИНГОВЫХ СИСТЕМАХ

**И.Н. ФОМИН**

аспирант кафедры прикладной информатики и программной инженерии,  
Международный факультет прикладных информационных технологий,  
Саратовский государственный технический университет им. Ю.А. Гагарина  
Адрес: 410054, г. Саратов, ул. Политехническая, д. 77  
E-mail: ignik16@yandex.ru

**Н.В. СЕРДЮКОВА**

аспирант кафедры прикладной информатики и программной инженерии,  
Международный факультет прикладных информационных технологий,  
Саратовский государственный технический университет им. Ю.А. Гагарина  
Адрес: 410054, г. Саратов, ул. Политехническая, д. 77  
E-mail: serdukova@orgcentr.com

В статье рассматривается задача построения расчетной модели измерения электроэнергии (*calculated measurement model of electric power*) в информационных системах для автоматизированного ценообразования на розничном рынке электроэнергии РФ. Объект исследования – розничный рынок электроэнергии (*the retail electricity market*) и бизнес-процессы энергосбытовых предприятий, предмет исследования – массивы данных об энергопотреблении различного рода потребителей (*consumption habits*).

Цель исследования – построение концепции и модели информационной системы, позволяющей рассчитывать розничную цену и стоимость потребленной электроэнергии.

Решенными задачами в исследовании явились:

- ♦ определение сущностей, которые являются параметрами для расчета цены и стоимости электроэнергии;
- ♦ классификация и определение критериев классификации учетных (*user data*) и расчетных показателей (*estimates data*), участвующих в автоматизированном ценообразовании;
- ♦ подбор различных вариантов интеграции этих параметров;
- ♦ моделирование структуры базы данных, которая формирует расчетную модель измерений электроэнергии.

В статье даны определения учетных и расчетных показателей, которые являются элементами отраслевых корпоративных информационных систем, описаны принципы формирования модели измерения электроэнергии (*calculated measurement model of electric power*) при создании и внедрении биллинговых информационных систем (*meter to cash systems*), а также даны определения различных терминов электроэнергетики с точки зрения информатики и информационных технологий.

Ценообразование на розничном рынке электроэнергии рассмотрено как некий социально-экономический процесс, к которому были применены методы математического моделирования и системного анализа. Предложенные критерии классификации и правила построения связей в структуре базы данных обладают устойчивыми связями элементов и организацией, что проявилось в выявленном снижении энтропии системы в целом по сравнению с системообразующими элементами и факторами.

Выявленные и описанные связи могут помочь на практике в задачах отраслевой программной инженерии и параметрической адаптации при разработке биллинговых информационных систем.

В описанном исследовании произведено моделирование структуры базы данных, формирующей расчетную модель измерений электроэнергии, позволяющую оперировать различными договорными и техническими параметрами электроснабжения (*contractual input*), а также определять розничную цену и стоимость потребленной электроэнергии.

Теоретические и аналитические исследования позволили создать авторам инновационный продукт в виде программы для ЭВМ «АС Энерго.УПЭЛ», зарегистрированной Свидетельством о госрегистрации № 2013615624 (опубликовано 17.06.2013), который успешно апробировался на предприятиях энергетической отрасли в нескольких регионах РФ.

**Ключевые слова:** ценообразование, розничный рынок электроэнергии, расчетная модель измерений электроэнергии, учетные показатели, расчетные показатели, информационные биллинговые системы, энергосбытовые компании.

## 1. Введение

**В**ажным условием достижения экономической и энергетической эффективности в электроэнергетике является степень внедрения информационных технологий. Энергосбытовая деятельность (retail power agency) – одна из тех областей, где информационные технологии (ИТ) играют ключевую роль. Оптимальная структура ИТ способна сделать компанию конкурентоспособной и эффективной, и наоборот, неправильно структурированные ИТ могут стать тормозом на пути ее развития [1].

Основой автоматизации производственных процессов в электроэнергетике является обработка и хранение данных об энергопотреблении. Над этой задачей работали американский ученый Ральф Кимбалл (Ralph Kimball), который определил практические приемы построения хранилищ данных и ключевые концепции многомерных моделей данных (multidimensional data model), используемых при хранении данных об объемах энергопотребления [2], и немецкий исследователь Харальд Орламюндер (Harald Orlamünder), который построил новую ролевую модель участников розничного рынка электроэнергии [3], которая может служить основой для разработки информационных систем с функциями автоматизированного ценообразования.

Корректное построение структуры биллинговых систем заключается в построении структуры хранения различных параметров учета энергопотребления в моделях, позволяющих оперировать различными договорными и техническими параметрами электроснабжения для периодических расчетов цены и стоимости потребленной электроэнергии на розничном рынке электроэнергии (РРЭ). В периодической отраслевой литературе имеется описание опыта автоматизации энергосбытовой деятельности компаний регионального и городского уровня, таких как МОЭК, «Волгоградэнергосбыт», «Мосэнергосбыт» и других предприятий, где использовались различные зарубежные программно-аппаратные платформы (SAP, Oracle E-Business Suite, Oracle CC&B). Концепции перечисленных СУБД (DBMS) включают в себя клиентское аналитическое приложение и набор инструментов, предназначенных для извлечения, анализа и представления информации из хранилища данных для нужд специалистов предприятия. Такого рода инструменты, как правило, не имеют прикладной ориентации и требуют от специалистов деятельность, связанную с построением запросов к хранилищу данных и визуализацией текстовых (linguistic evidence) и числовых данных (numerical data), при которой используются многомерные модели данных.

К сожалению, коммерческие организации не публикуют результаты своих исследований, в том числе в научных журналах. Авторы статьи участвовали в постановке технических задач и внедрении отечественных информационных биллинговых систем российских разработчиков в компаниях «Арктик–энерго» (Мурманская область), «Оборонэнергосбыт» (Москва), во множестве региональных энергосбытовых организаций Волгограда, Владивостока, Ярославля, Самарской и Саратовской областей. Анализируя результаты проектов, можно сделать вывод, что применение в энергосбытовой деятельности биллинговых информационных систем дает значительные положительные эффекты как оперативного, так и стратегического характера на долгосрочную перспективу, в интересах субъектов энергетического рынка. При этом информационные биллинговые системы для энергосбытовых организаций (retail power enterprises) на розничном рынке электроэнергии строятся интуитивно, концепции построения отраслевых информационных систем и структур баз данных в литературе не описаны.

## 2. Модели измерения электроэнергии

Построение модели информационной системы, позволяющей определять розничную цену и стоимость потребленной электроэнергии, оперируя различными договорными и техническими параметрами электроснабжения, предполагает параметрическую адаптацию программного обеспечения (ПО). В процессе электроснабжения (electric power supply) потребителей происходят различные технические (specifications subject to change) и договорные изменения, которые, как следствие, приводят к изменениям параметров расчетной модели. При этом параметрическая адаптация подразумевает изменение переменных, определяющих поведение программы. При таком подходе можно менять параметры или заставить приложение следовать какой-либо иной из имеющихся стратегий [4]. Параметры адаптации приложений или информационных систем определяются для каждой точки поставки (supply point) электроэнергии.

**Расчетная модель измерения электроэнергии (РМИ)** (calculated measurement model of electric power, СММ) – это модель, описывающая (или отражающая в информационной системе) правила и методы обработки информации о ценах и стоимости потребляемой электроэнергии. Расчетная модель измерения отражает экономические условия электроснабжения и определяет структуру базы данных, в которой хранятся данные, необходимые

для расчета цен и стоимости потребленной энергии. Задачей расчетной модели измерений является формирование корректных учетных и расчетных показателей [5].

**Учетные показатели** — это количественные величины, которые требуется хранить в базе данных информационной системы для проведения расчетов или анализа данных. Эти данные снимаются с приборов учета электроэнергии, вводятся вручную, загружаются из внешних систем, реже — находятся расчетным путем. В общем случае, учетными показателями можно назвать значения функций с определенным набором аргументов, которые в дальнейшем участвуют в каких-либо расчетах или используются в отчетах информационной системы. С помощью учетных показателей реализуются учетные функции специализированного программного обеспечения.

**Расчетные показатели** — это различные показатели, сформированные в автоматизированной системе в результате применения расчетных функций [6].

Учетные и расчетные показатели в базах данных отраслевых информационных систем хранятся в виде измерений. Согласно концепции Ральфа Кимбалла, измерение — это интерпретация факта (facts for measurements) с некоторой точки зрения в реальном мире, содержащая учетные и расчетные показатели, которые связаны по смыслу между собой. Обычно измерения представляются как оси многомерного пространства, точками которого связывают определенные факты реального мира с числовыми или лингвистическими показателями, характеризующими эти факты. В отраслевых информационных системах эти лингвистические показатели являются разрезами аналитического учета (incisions analytical accounting). В многомерной модели каждый учетный или расчетный показатель связан с одной или несколькими осями — измерениями. Измерения — это ключевая концепция многомерных баз данных. Многомерное моделирование (dimensional modeling (DM) предусматривает использование измерений для предоставления максимально возможного контекста для измеряемых фактов (facts for measurements) [2]. Хранение данных, для целей расчета стоимости, должно производиться согласно сформированной в информационной системе расчетной модели измерений.

Используя теорию множеств и теорию графов, расчетную модель измерений можно задать как конечное множество  $Y$ , состоящее из  $n$  элементов, определяющих расчетные показатели для  $j$ -той точки учета

$$Y_j = \{1, 2, \dots, n\},$$

которые и формируют вершины ориентированного графа, и множество дуг (set of edges) подмножества

(subset)  $P = \{p_1, p_2, \dots, p_n\}$ , при этом  $P \subseteq Y$ . В реальных информационных системах базы данных формируют набор таблиц, которые отражают данные по составляющим, используемым при расчете цен и стоимости. Количество и иерархия элементов определяется расчетной схемой для потребителя, а также задачами ценообразования (формирования расчетной модели измерения). Расчетная модель измерения электроэнергии в биллинговых информационных системах формируется для целей определения цены и стоимости, путем связи объектов или справочников и может отображаться в древовидной форме. Элементами этих множеств являются множества, содержащие данные об объектах электроснабжения (power supply objects) потребителей и связях между ними ( $P^O$ ), параметрах расчета объемов ( $P^V$ ), и данные о параметрах расчета цен и стоимости по объектам электроснабжения ( $P^C$ ). Таким образом,  $P = \{P^O, P^V, P^C\}$ . Такое представление дает возможность использовать теорию множеств для классификации учетных и расчетных показателей и разработки структуры базы данных информационных биллинговых систем.

### 3. Классификация учетных и расчетных показателей

Оперируя учетными и расчетными показателями можно автоматизировать процесс ценообразования на розничном рынке электроэнергии.

**Ценообразование в автоматизированной биллинговой системе** — это автоматический или полуавтоматический расчет цены на электроэнергию с использованием динамически изменяющихся учетных или расчетных показателей.

Основными задачами процесса автоматизированного ценообразования являются:

- ◆ определение различных учетных показателей, участвующих в формировании цены и отслеживание изменений этих показателей;
- ◆ ввод и хранение данных об учетных показателях согласно модели измерения электроэнергии и отслеживание изменений модели;
- ◆ получение новых расчетных показателей, участвующих в расчете цен и стоимости электроэнергии.

В процессе исследования для поиска технических решений описанных выше задач к каждой группе потребителей электроэнергии (каждому объекту электроснабжения на РПЭ) был подобран набор атрибутов, необходимых для выполнения расчетных операций. Выделяют четыре группы потребителей: базовые потребители, население, прочие потребители, организации, оказывающие услуги по пере-

даче электроэнергии и приобретающие ее в целях компенсации потерь [7]. Основными критериями, по которым подвергались классификации атрибуты объектов, расчетные и учетные показатели, стали:

- ◆ конечный набор объектов и классов;
- ◆ возможность хранения показателей, характеризующих объекты, в реляционных базах данных;
- ◆ соответствие отраслевому законодательству.

Также при классификации были выделены параметры и атрибуты, определяющие цену для каждого объекта учета, метод ее расчета и условия использования данной цены при расчете стоимости. Упомянутая выше неоднозначность использования элементов РМИ при классификации сущностей для целей ценообразования предполагает возможность использования этих элементов как объектов либо как классов в зависимости от степени их характерности.

В связи с различием в ценообразовании на РРЭ в ценовых (ЦЗ) и неценовых зонах (НЦЗ) был произведен анализ факторов, определяющих особенности формирования цен и стоимости указанных выше групп потребителей в различных зонах оптового рынка и их степени доминирования в реальных договорных условиях электроснабжения. В результате было выявлено конечное количество наборов технико-экономических условий потребления электроэнергии, позволяющих применять тот или иной вид ценообразования и способ расчета стоимости потребленной электроэнергии.

Анализ проводился с применением метода экспертной классификации, принцип которого состоит в обоснованном переносе классификационных решений, полученных для одного объекта, на некоторые другие объекты. Базой знаний для обоснования и переноса служит отношение доминирования, которое, в свою очередь, строится из отношений характерности [8].

При исследовании способов расчета цен и стоимости электроэнергии в НЦЗ был получен контекст  $K_1 = (G_1, P_1, I_1)$ , где  $G_1$  – группы потребителей, соответствующие различным способам расчета цены и стоимости в НЦЗ,  $P_1$  – множество признаков, по которым классифицировались элементы множества  $G_1$ , а  $I_1$  – множество значений этих признаков.

При исследовании способов расчета цен и стоимости потребителей в ЦЗ был получен контекст  $K_2 = (G_2, P_2, I_2)$ , где  $G_2$  – группы потребителей, соответствующие различным способам расчета цены и стоимости в ЦЗ,  $P_2$  – множество признаков, по которым классифицировались элементы множества  $G_2$ , а  $I_2$  – множество значений этих признаков.

Задача определения наиболее общих призна-

ков, определяющих способ расчета цены и стоимости, решалась методом совмещения контекстов  $K_1 = (G_1, P_1, I_1)$  и  $K_2 = (G_2, P_2, I_2)$ , в результате которого было получена решетка  $C(K)$  контекста  $K = (G, P, I)$ , представляющая собой множество различных наборов технико-экономических условий потребления электроэнергии, позволяющих применять тот или иной вид ценообразования и способ расчета стоимости потребленной электроэнергии.

В контексте  $G$  – группы потребителей, соответствующие различным способам расчета цены и стоимости,  $P$  – множество признаков (характеристик), по которым классифицируются элементы множества  $G$ , соответственно  $I$  – множество значений этих признаков.

Полученные наборы  $C(K)$  зависят от административных и договорных условий потребления, а также от характеристик применяемых приборов учета. Эти наборы влияют на процесс ценообразования и в законодательстве РФ называются «ценовая категория».

Ценовая категория – это набор технико-экономических условий потребителя, позволяющий применять тот или иной вид ценообразования и способ расчета стоимости потребленной электроэнергии. Энергосбытовые компании в статусе гарантирующего поставщика вынуждены административно относить потребителя к одной из ценовых категорий, а независимые энергосбытовые компании технически относят потребителя к одной из ценовых категорий.

В полученном контексте каждому элементу  $g$  множества  $G$  соответствует множество признаков  $P$  и множество значений этих признаков  $I_g = \varphi(g)$ , соответствующих той или иной ценовой категории.

Используемый выше контекстный подход позволил выделить различные множества характеристик, участвующие в формировании реляционной модели базы данных информационной биллинговой системы. Для этого ценовая категория была выражена как конечное множество, состоящее из элементов, определяющих характеристики для  $j$ -го объекта электроснабжения:

$$K_i^C = \{p_{n1}, p_{n2}, \dots, p_{in}\}, K_{ij}^C \subset P^C,$$

где  $p_{jk}$  –  $k$ -ая характеристика, определяющая ценовую категорию для  $j$ -го объекта электроснабжения,  $n = 8$  – количество характеристик, среди которых можно выделить следующие:

1. отношение потребителей к ценовым или не ценовым зонам;
2. категория потребителей (население, крупные, мелкие потребители, сетевые организации);

3. наличие факта (требования) трансляции цен электроэнергии и мощности ОРЭМ;

4. тип тарифа на энергию (одноставочный или двухставочный);

5. тип тарифа на услуги по передаче электроэнергии (одноставочный или двухставочный);

6. способ учета объемов (интегральный в целом за период, интервальный по часам, по зонам суток);

7. число зон – для способа учета объемов по зонам суток;

8. наличие факта (требования) почасового планирования.

Множество ценовых категорий  $K^C = \{K_i^C\}$  определяется различными комбинациями значений характеристик, определяющих ценовую категорию, но в целом данное множество конечно и содержит различные варианты ценовых категорий для ценовых и неценовых зон.

В России цены на электрическую энергию (мощность), поставляемую потребителям, не могут быть выше некоторых ее предельных уровней. В концепции, предложенной авторами настоящей публикации, предельный уровень цен в автоматизированной системе – это расчетный показатель, полученный в соответствии с порядком формирования цены, определяемым связанной с ним ценовой категорией и хранящийся в реляционной базе данных (БД) или в соответствующем регистре БД для использования в определении стоимости потребленной электроэнергии.

Расчет предельных уровней цен можно выразить как некую функцию, параметрами которой является набор показателей ХС, используемых при формировании предельного уровня цен:

$$C = f(X^C)$$

Набор элементов множества  $X^C$  различен для ценовых и неценовых зон или иными словами зависит от ценовой категории, то есть  $X^C = f(K^C)$ .

Для ЦЗ элементами множества  $X^C$  являются такие учетные показатели как средневзвешенные нерегулируемые цены ОРЭМ (СВНЦ) – ( $D_i^{ATS}$ ), тарифы на передачу ( $T_i^{GRID}$ ), тарифы на услуги инфраструктуры ОРЭМ ( $T_i^{INF}$ ), сбытовые надбавки ( $\Delta_i^C$ ):

$$X_i^{C\_pa} = \{D_i^{ATS}, T_i^{GRID}, T_i^{INF}, \Delta_i^C\}.$$

Для НЦЗ элементами множества  $X^C$  являются такие учетные показатели как: данные коммерческого оператора, влияющие на формирование розничных цен  $D_i^{KO}$  регулируемые розничные тарифы (regulated prices  $T_i^{REG}$ ), удельная стоимость электроэнергии на ОРЭМ  $J_i$ :

$$X_i^{C\_npa} = \{D_i^{ATS}, T_i^{REG}, J_i\}.$$

Данные учетные показатели, входящие в состав  $X_i^C$  изменяются в зависимости от того с какой группы точек поставки (ГТП) ( $r_1$ ), каким субъектом ОРЭМ (энергосбытовой компанией или гарантирующим поставщиком) ( $r_2$ ), по какому типу договора ( $r_3$ ) осуществляется поставка электроэнергии (договор электроснабжения или договор купли-продажи). Также, учетными показателями являются сетевая организация, по сетям которой осуществляется электроснабжение ( $r_4$ ), в каком регионе ( $r_5$ ) и по какому уровню напряжению ( $r_6$ ) производит передачу электроэнергии, к какой ценовой категории ( $r_7$ ) и группе потребителей ( $r_8$ ) относится объект электроснабжения. Кроме того, имеет значение такой показатель как зона суток (TOU) ( $r_9$ ), для которой рассчитывается цены. Таким образом, эти наборы учетных показателей являются аргументами функции  $X_i^C = f(r_i | i = 1..n)$ .

Это множество представляет собой объединение нескольких множеств, определяющих структуру хранения данных о каждом из элементов множеств ХСцз, ХСнцз, которые могут соответствовать разрезам аналитического учета в информационных биллинговых системах.

Совокупность учетных показателей  $\{r_i | i = 1..n\}$ , от которых зависят данные о составляющих, входящих в состав предельных уровней цен, можно представить как некое конечное множество  $R = \{r_i | i = 1..n\}$ , состав которого зависит от ценовой категории, т.е.

$$R = f_c(K^C), R \subset X.$$

В свою очередь, каждый из элементов множества  $X$  входит в состав других множеств, представляющих собой данные о том или ином виде показателя, используемого при расчете предельных уровней цен.

$$D_i^{ATS} \subset D^{KO}, T_i^{GRID} \subset T^{GRID}, \Delta_i^C \subset \Delta^C, T_i^{REG} \subset T^{REG}.$$

Анализ законодательства и массивов данных о ценах различных энергосбытов или гарантирующих поставщиков показал, что данные о предельных уровнях цен зависят от множества учетных показателей  $R$ , определяющих их составляющие, которые участвуют в расчете этих цен. Кроме того, эти же учетные показатели, от которых зависит формирование предельных уровней цен, являются свойствами объектов электроснабжения, позволяющими использовать, при расчете стоимости, сформированные предельные уровни цен.

Совокупность учетных показателей, в разрезе которых рассчитываются предельные уровни цен,

определяет множество различных вариантов расчета предельных уровней цен.

Вариант расчета предельного уровня цен – это элемент расчетной модели, относящий использование одной и той же ценовой категории и определяющей метод расчета цен или предельных уровней цен с различными учетными показателями для схожих по набору свойств объектов электроснабжения.

Исходя из множества вариантов расчета цен и предельных уровней цен (дифференциация по уровням напряжения, зонам суток, группам потребителей, типам договоров, ГТП), можно объединить некоторые показатели, одинаковые для различных ГТП участников ОРЭМ, в так называемые тарифные группы и определить взаимозависимость учетных показателей. Говоря языком теории множеств можно выделить в множестве  $R$  подмножество  $G^T = \{t_1, t_2, \dots, t_n\}$ , где  $G^T \subset R$ . Элементами множества  $G^T$  являются такие учетные показатели как ценовая категория, уровень напряжения, вид потребителя, тип договора. Очевидно, что при этом уменьшится количество показателей  $\{r_i | i = 1..n\}$ , от которых зависят данные о составляющих предельных уровней цен и данные о предельных уровнях цен, поскольку вместо элементов, входящих в множество  $G^T$ , можно использовать один элемент, определяющий данный набор.

Тарифная группа ( $G^T$ ) – это совокупность технико-экономических условий потребителя, позволяющая определять тот или иной вариант расчета цен или предельных уровней цен для той или иной ценовой категории потребителя. Тарифная группа в данной концепции, включает в себя ценовую категорию, задающую метод расчета цены и стоимости для ГТП либо совокупности ГТП, к которой относятся потребители данной тарифной группы. Эти совокупности являются основными показателями, определяющими параметры расчета цен и стоимости для объектов электроснабжения на РРЭ, находящиеся в данной тарифной группе и в данной ГТП.

Соответственно, для  $j$ -го объекта электроснабжения можно выделить еще одно свойство  $G_j^T$ , определяющее отношение данного объекта к тарифной группе, причем  $G_j^T \in G^T$ . Поскольку тарифная группа включает в себя ценовую категорию ( $K^C \subset G^T$ ), то можно утверждать, что отношение объекта электроснабжения к тарифной группе определяет также и его отношение к ценовой категории, либо определяет способ расчета цены и стоимости для данного объекта.

Расчет стоимости для  $j$ -го объекта электроснабжения, а также расчет цен и расчет уровней цен

можно выразить как некую функцию  $f_s$ , параметрами которой являются наборы данных о ценах  $C_j$  и объемах энергопотребления  $V_j$ , которые, в свою очередь, зависят от соответствующих данному объекту электроснабжения параметров расчета цены  $P_j^C$  и параметров расчета объемов  $P_j^V$ .

$$S_j = f_s(C_j, V_j), C_j = f_c(X_j^C) = f_c(f_{X_j}(P_j^C)), V_j = f_v(P_j^V).$$

На основе проведенного исследования можно утверждать, что набор данных о характеристиках, определяющих алгоритм расчета цены и стоимости  $P_s$  для данного объекта электроснабжения, должен содержать такие учетные показатели как субъект ОРЭМ (гарантирующий поставщик или энергосбытовая организация, по ценам которой производится расчет стоимости), ГТП субъекта ОРЭМ, тарифная группа и сетевая организация, чьи тарифы применяются при определении стоимости услуг по транспорту электроэнергии.

#### 4. Моделирование базы данных информационной системы

При моделировании базы данных с помощью ER-диаграмм данное множество  $P_s$  было представлено в виде ассоциативной сущности «Параметры расчета цен и стоимости объектов электроснабжения на РРЭ», определяющей связи между тремя различными объектами информационной системы: «Объекты электроснабжения», «ГТП» и «Тарифные группы». Данная связь является базой, определяющей, по каким тарифным группам и для каких ГТП субъектов ОРЭМ следует рассчитывать цены и предельные уровни цен, а также определяет, предельные уровни какой тарифной группы и какого субъекта ОРЭМ следует использовать при расчете стоимости для конкретного объекта электроснабжения. Представленная на рис. 1 модель отражает основные учетные показатели, определяющие параметры расчета цен и стоимости для объектов электроснабжения, их атрибуты, а также устанавливает взаимосвязи между этими показателями.

В концепции реляционной модели данных, изображенный на рис. 1 элемент расчетной модели для целей автоматизированного ценообразования отражает основные таблицы, связанные с формированием цен для объектов электроснабжения, их структуру (поля) и связи между таблицами. Таблица ПАРАМЕТРЫ\_РАСЧЕТА (calculation\_Options) в модели отражает взаимосвязи между различными учетными и расчетными показателями, описываемыми множеством  $P^C$ . Таблицы, объединенные в группы «Составляющие ПУЦ для ЦЗ» и «Состав-

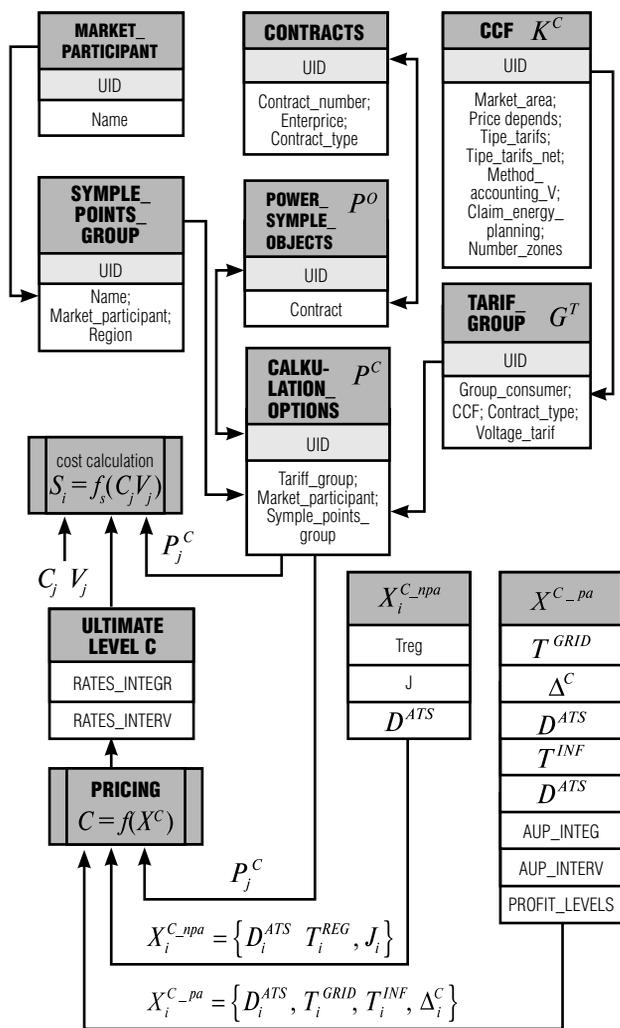


Рис. 1. Схема расчетной модели измерения электроэнергии для задач автоматизированного ценообразования

ляющие ПУЦ для НЦЗ», представляют собой описанные выше элементы множеств  $X^{C-ppa}$  и  $X^{C-ppa}$  соответственно. Множество  $R$  в базах данных реальных информационных систем является одной из осей измерений и используется как разрезы аналитического учета, определяя структуру хранения данных, используемых для расчета цен и предельных уровней цен.

Получение данных этих таблиц при расчете предельных уровней цен может быть организовано посредством соответствующих запросов, которые можно выразить функцией  $f_X(P_j^C)$ .

### 5. Заключение

Предложенная модель классификации учетных и расчетных показателей, участвующих в процессе автоматизированного ценообразования, наглядно показывает, в каких разрезах следует хранить показатели, участвующие в расчете цен, а также чем могут определяться параметры расчета цен и стоимости для объектов электроснабжения потребителя.

Моделирование структуры базы данных, формирующей расчетную модель измерений электроэнергии, может быть использовано при проектировании информационных систем, используемых для автоматизации деятельности энергосбытовых организаций. Результаты исследования также могут быть использованы и в разработке организационной и технической концепции подбора оптимальных договорных условий энергоснабжения, что напрямую влияет на энергоэффективность отрасли. ■

### Литература

1. Чернов С.С. Энергосбытовая деятельность в условиях реформирования: проблемы и перспективы // Проблемы современной экономики. 2011. № 4 (40). С. 157-164.
2. Kimball R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses (Практические приемы построения хранилищ данных). New York: John Wiley & Sons, 1996.
3. Orlamünder H. Der Einsatz von Informations- und Kommunikationstechnik in Stromnetzen – ein Nachhaltiges Energieinformationsnetz (Использование информационно-коммуникационных технологий в электрических сетях – устойчивая информационная энергетическая сеть). Stuttgart: Alcatel-Lucent, 2009.
4. Шульга Т.Э. Метод построения восстанавливающих последовательностей для систем без потери информации // Системы управления и информационные технологии. 2009. № 1.3 (35). С. 407-411.
5. Фомин И.Н. Модели измерений в информационных системах энергосбытовых организаций // Высокие технологии, исследования, финансы. СПб.: Изд. Политехнического ун-та. 2013. С. 213-217.
6. Фомин И.Н. Методика поэтапного автоматизированного расчета стоимости потребленной электроэнергии // Труды Академэнерго. 2013. № 3. С. 103-115.
7. Богачкова Л.Ю., Иншаков О.В., Москвичев Е.А. Совершенствование функционирования распределительных сетевых компаний на либерализуемых розничных рынках электроэнергии в РФ / Препринт. Волгоград: Изд. ВолГУ, 2012.
8. Соловьев С.Ю., Стельмашенко Д. Е. Применение принципов экспертной классификации для анализа формальных понятий // Бизнес-информатика. 2013. № 4(26). С. 53-57.

# COMPUTATIONAL MODEL FOR MEASURING ELECTRIC POWER IN METER-TO-CASH SYSTEMS

**Igor FOMIN**

*Post-graduate Student, Department of Applied Informatics and Software Engineering,  
International School of Applied Information Technology,  
Yuri Gagarin State Technical University of Saratov*

*Address: 77, Politechnicheskaya Street, Saratov, 410054, Russian Federation*

*E-mail: ignik16@yandex.ru*

**Nadezhda SERDYUKOVA**

*MSc Program Student, Department of Applied Informatics and Software Engineering,  
International School of Applied Information Technology,  
Yuri Gagarin State Technical University of Saratov*

*Address: 77, Politechnicheskaya Street, Saratov, 410054, Russian Federation*

*E-mail: serdukova@orgcentr.com*

*This article addresses building a computational model for measuring electric power in information systems designed to support automated pricing in the retail electricity market of the Russian Federation.*

*The object of this study is the retail electricity market and business processes of retail power supply enterprises. The subject of the study is arrays of energy consumption habits data.*

*The purpose of this study is to build a conceptual framework and an information system model enabling to calculate retail prices and costs of electric power consumed.*

*This study has solved the following tasks:*

- ◆ *to identify entities that constitute parameters to calculate electric power prices and costs;*
- ◆ *to classify user data and estimate data involved in automated pricing, and to formulate classification criteria;*
- ◆ *to select various options to integrate these parameters;*
- ◆ *to simulate the structure of a database that forms the computational model for measuring electric power.*

*This paper presents definitions of user data and estimate indicators, which are elements of industry-specific corporate information systems, describes underlying principles applied to build the computational model for measuring electric power when designing and implementing meter-to-cash systems, as well as formulates definitions of various electric power concepts from the perspective of information technologies.*

*Pricing in the retail electricity market has been regarded as a socio-economic process, towards which mathematical simulation and systems analysis methods have been applied. The suggested classification criteria and rules for links building in the database structure enable to ensure sustainable element links and organization that has manifested itself in revealed decrease in entropy of the system as a whole compared to the backbone elements and factors.*

*The detected and described links can offer practical help to address industry-specific software engineering and parametric adaptation challenges faced in development of meter-to-cash systems.*

*This study involves simulation of the database structure that forms the computational model for measuring electric power enabling to operate various power supply contractual and technical inputs, as well as to determine retail prices and cost of electric power consumed.*

*The theoretical and analytical research enabled the authors to create an innovative product – a software solution «AS Energo. UPEL» (certificate of state registration # 2013615624 issued on 17.06.2013) that has been successfully tested at power supply enterprises in several Russian regions.*

**Key words:** pricing, electricity retail market, computational model for measuring electric power, user data, estimate data, meter-to-cash systems, retail power supply companies.



# АВТОМАТИЗАЦИЯ ПРОЦЕССА ОБУЧЕНИЯ

## **В.Г. ЧЕБОТАРЁВ**

кандидат технических наук, доцент кафедры моделирования и оптимизации бизнес-процессов, факультет бизнес-информатики, Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: vchebotarev@hse.ru

## **А.И. ГРОМОВ**

кандидат химических наук, профессор кафедры моделирования и оптимизации бизнес-процессов, факультет бизнес-информатики, Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: agromov@hse.ru

Материалы статьи основаны на результатах, полученных участниками (преподавателями и студентами) научно-исследовательского семинара факультета бизнес-информатики НИУ ВШЭ. Цель — исследование возможностей управления обучением как процессом, применение современных информационных технологий и инструментов. Дано определение процесса обучения как совместной деятельности участников процесса в организации усвоения учебной дисциплины, получения субъективно новых или новых предметных и процедурных знаний. В качестве предмета исследования выбран процесс обучения конкретной дисциплине «Моделирование и оптимизация бизнес-процессов». Выполнен анализ особенностей процесса обучения, определены требования к управлению процессом обучения, выполнено моделирование и оценка возможностей автоматизированного управления процессом обучения. Для анализа использована информация об организации и проведении реальных занятий на факультете бизнес-информатики. Выявлены следующие требования к процессу обучения: быстрая адаптация к изменениям, гибкое управление с учетом различных методов обучения, постоянное взаимодействие участников процесса обучения, а также создание возможностей для творчества, рефлексии, для использования субъектности участников процесса. Для моделирования использовались различные нотации: eEPC и диаграммы взаимодействия и поведения субъектов. Первая нотация относится к классической методологии моделирования ARIS, а вторая — к субъектно-ориентированной методологии управления процессами Metasonic. Показано, что удовлетворение всех требований к процессу обучения обеспечивается при субъектно-ориентированном подходе и практически невозможно при классическом подходе. Определены принципы разделения обобщенного процесса обучения на мелкие процессы, разработаны их модели в виде диаграмм взаимодействия и диаграмм поведения субъектов. Представлены возможности изменения процесса обучения участниками процесса без программирования, с учетом автоматизации генерации приложения в Metasonic S-BPM Suite. Результаты исследования могут быть полезны для любых организаций, задумывающихся о переходе от традиционной жесткой структуры к более современной рефлексивной организации с сетевым взаимодействием сотрудников.

**Ключевые слова:** процесс обучения, моделирование процесса, автоматизация процесса, субъектно-ориентированное управление процессом, S-BPM, гибкий процесс, рефлексивность.

### Введение

По мнению ряда исследователей, суть обучения постепенно трансформируется от простого получения «предметных знаний» к получению «процедурных знаний» [1]. Рассма-

триваются три вида обучения [2]: 1) простейшее обучение тому, что находится на поверхности, с целью приспособиться к тому, что является общепринятым; 2) критическое переосмысление гипотез, изучение контекста деятельности и возмож-

ные коренные изменения собственных целей; 3) осмысление и переосмысление самого процесса, самокритика, осознание и исправление мыслительных моделей, проверка отправных смысловых точек действия. Второй и особенно третий виды обучения связаны с получением «процедурных знаний», с развитием рефлексии у обучающихся и обучающихся.

### 1. Требования к процессу обучения

В литературе часто используется термин «учебный процесс», а в образовательных стандартах (ФГОС) упоминается «образовательный процесс». Оба термина недостаточно однозначны, поэтому будем использовать более точный и однозначный термин — «процесс обучения». Отметим, что исследование не касается *дидактической стороны* обучения, а посвящено *организационной стороне* обучения. Процесс обучения рассматривается как один из процессов организации, а управление процессом состоит из моделирования, анализа, совершенствования, автоматизации и контроля.

Процесс обучения в вузе отличается рядом особенностей, затрудняющих управление. Формально обучение в вузе строго регламентируется стандартами, учебными планами и программами. Реально обучение требует постоянных перемен и творческого подхода. Выделим две категории причин перемен: 1) переменны в обществе, научные исследования, состояние экономики и т.п.; 2) уникальность аудитории и индивидуумов, противоречия во взаимоотношениях с бюрократическими структурами, обратные связи, внезапные события и т.п. Все более важной становится научно-исследовательская деятельность, которая включает как научные исследования преподавателей, так и обучение студентов на семинарах, в проектах и т.д. Стремление к получению процедурных знаний — первый шаг к научным исследованиям.

Определим **процесс обучения** как *совместную деятельность участников процесса по организации усвоения учебных дисциплин для получения предметных и процедурных знаний*. Термины «субъект» и «субъектность» определим следующим образом [3]: «субъект» — носитель деятельности, сознания, познания и рефлексии; «субъектность» — характеристика субъекта (как носителя деятельности, сознания, познания и рефлексии), отражающая полноту его актуальных и потенциальных воз-

можностей и степень их реализации в настоящем и будущем. Под «обобщенным процессом» обучения будем понимать типовой процесс, содержащий общие черты, схемы, принципы обучения по множеству дисциплин. Под метамоделью процесса обучения будем понимать описание структуры обобщенного процесса, определения взаимодействия элементов его структуры. Совместная деятельность участников процесса по получению знаний предполагает возможность использования субъектности любого участника процесса, независимо от его роли.

Как происходит управление процессом обучения? Для управления часто используют следующие методы: 1) непосредственные управляющие воздействия обучающего; 2) опосредованные воздействия «обобщенного преподавателя»; 3) самоуправление обучаемого. К недостаткам методов относятся: отсутствие взаимодействия с обучаемыми, ориентация на несуществующего усредненного обучаемого и одностороннее обучение, отсутствие постоянной обратной связи. Взгляд на обучение как на процесс позволяет утверждать, что возможности управления обучением напрямую зависят от возможностей автоматизации процесса. Обычно для автоматизации обучения используется классический подход, основанный на разработке иерархических автоматизированных систем, состоящих из подсистем, каждая из которых, в свою очередь, состоит из модулей. Авторам не известна ни одна из автоматизированных систем, которая предусматривала бы автоматизацию и управление обучением как процессом, все современные системы являются *обеспечивающими*. С учетом результатов ряда исследований [4, 5, 6], сформулируем следующие требования к процессу обучения:

1. Быстрая адаптация процесса к изменениям. Обеспечивается, например, созданием условий для самоорганизации системы.
2. Гибкое управление процессом, возможность применения различных методов. Обеспечивается самоорганизацией и децентрализацией.
3. Постоянное взаимодействие участников процесса, как предметное, так и процедурное. Приводит к аддитивному эффекту.
4. Создание возможностей для рефлексии, творчества, использования субъектности участников процесса. Расширяет использование интеллектуальных ресурсов участников.
5. Интеграция в процессе документов и обеспе-

чивающих информационных систем. Позволяет рационально использовать необходимые ресурсы.

б. Оценка эффективности выполнения каждого экземпляра. Повышает возможности по совершенствованию процесса.

Реализация этих требования при автоматизации процесса с помощью традиционных подходов (например, BPMN) чрезвычайно трудоемка. Однако при использовании субъектно-ориентированного подхода выполнение требований становится возможным.

## 2. Субъектно-ориентированный подход к управлению процессом обучения

Методология субъектно-ориентированного подхода реализована в инструментальной системе Metasonic S-BPM Suite для моделирования, валидации и немедленного исполнения бизнес-процессов. «Немедленное исполнение» означает, что преобразование моделей в исполняемое приложение (workflow) происходит автоматически при выгрузке моделей на сервер. В отчете за 2014 г. Gartner назвала Metasonic S-BPM Suite инновационным, эффективным и интригующим предложением на рынке BPM [7]. Применение S-BPM позволит не только существенно сократить расходы на автоматизацию и кардинально повысить адаптивность процессов, но и существенно продвинуться на пути к созданию рефлексивного предприятия XXI века, основанного на главенствующей роли субъектности при выполнении процессов [8].

Процесс в S-BPM выглядит как обмен сообщениями между участниками. Каждый участник рассматривается не как функциональный ресурс, а как субъект, обладающий интеллектом, способностями к творчеству и рефлексии. Представим, как выглядит реализация S-BPM в инструментальной системе Metasonic S-BPM Suite. В ее состав входят:

- ◆ Metasonic Build – среда проектирования процессов и ИТ-разработки;
- ◆ Metasonic Proof – среда верификации процессов;
- ◆ Metasonic Flow – среда исполнения и контроллинга процессных приложений;
- ◆ Metasonic Base – среда администрирования сервера и приложений.

Существует два основных типа процессных моделей: «модель взаимодействия» и «модель поведения». Первая модель описывает взаимодействие

субъектов между собой в виде обмена информационными сообщениями, а вторая отражает участие субъекта в процессе в виде описания его состояний и последовательных переходов из одного состояния в другое. При этом достаточно знания всего пяти элементов нотации: «субъект», «сообщение», «состояние отправления», «состояние получения» и «функциональное состояние». Последние три элемента отражают состояния субъекта при выполнении процесса. Интересно, что состояние ожидания, которое занимает большую часть жизни человека, может быть представлено в модели состояний как функциональное состояние. При моделировании процесса в этой же модели выполняется проектирование исполняемого приложения. Реализация (генерация кода) происходит автоматически при загрузке моделей в хранилище на сервере. Таким образом, кодирование приложения отсутствует, а простота нотаций позволяет привлечь к моделированию всех участников процесса. Субъектно-ориентированный подход дает возможность участникам процесса самостоятельно проектировать и быстро перепроектировать модели процесса, имитировать выполнение моделей для согласования (валидации) процесса с коллегами, используя для этого рефлексии, творческие способности, субъектность всех участников процесса, а также немедленно превращать модели процесса в исполняемые приложения путем загрузки этих моделей на сервер.

## 3. Моделирование процесса обучения в вузе

Прежде, чем перейти к моделированию процесса в среде Metasonic Build, выделим важные дополнительные особенности процесса обучения:

1. Как правило, схемы и принципы обучения (в вузе, на факультете, на кафедре) имеют много общего для разных дисциплин. Эти общие черты могут быть представлены в обобщенном процессе обучения, представленном в виде метамодели.
2. Обучение конкретным дисциплинам в вузе имеет много особенностей. Особенности могут быть отражены при детальном описании элементов обобщенного процесса (модели конкретных процессов).

В соответствии с принципами субъектно-ориентированного подхода, при моделировании используем направление «снизу-вверх». Сначала моделирование выполним для конкретной дисциплины «Моделирование и совершенствование

Таблица 1.

**Компоненты и субъекты процесса обучения дисциплине  
«Моделирование и совершенствование бизнес-процессов»**

№	Процесс	Цикл	Участники	Внутренние субъекты (бизнес-роли)	Внешние субъекты (бизнес-роли)
1	Теоретическое занятие (лекция)	1 неделя	Преподаватель-лектор, студенты курса	«Лектор» «Студент»	«Методист» «Менеджер»
2	Групповое практическое занятие (семинар)	1 неделя	Преподаватель семинара, студенты группы	«Преподаватель» «Семинарист»	«Методист» «Менеджер»
3	Внеаудиторная (домашняя) работа	1-7 дней	Студенты курса	«Семинарист»	Нет
4	Экзамен (тестовое задание)	8-9 недель	(Преподаватель-лектор, студенты курса) или (Преподаватель семинара, студенты группы)	«Экзаменатор» «Экзаменуемый»	«Методист» «Менеджер»
5	Подведение итогов	1 неделя	Преподаватель, менеджер учебной части	«Методист» «Менеджер»	«Лектор» «Преподаватель» «Экзаменатор»

бизнес-процессов», затем полученные результаты обобщим для других дисциплин кафедры и сформируем метамодель обобщенного процесса. Обучение по дисциплине может быть представлено в виде пяти простых процессов (табл. 1), отличающихся разными циклами, некоторой самостоятельностью и небольшим числом участников (субъектов, бизнес-ролей). Например, при выполнении процесса «Теоретическое занятие (лекция)» происходит обмен сообщениями (рис. 1) между внутренними субъектами процесса «Лектор» и «Студент» (множественным). Внутренними являются субъекты, участвующие в одном и том же процессе. Кроме того, «Лектор» получает информацию о студентах от внешнего субъекта «Менеджер учебной части» и передает оценки для подведения итогов внешнему

субъекту «Методист». Обмен сообщениями между внутренними субъектами процесса отражает выполнение этого процесса, а обмен сообщениями с внешними субъектами отражает интерфейс между разными процессами.

Для описания взаимодействия субъектов используется диаграмма взаимодействия. Аналогично могут быть представлены взаимодействия внутренних и внешних субъектов в процессах «Групповое практическое занятие (семинар)», «Внеаудиторная (домашняя) работа», «Экзамен (тестовое задание)» и «Подведение итогов». Процесс «Подведение итогов» интегрирует остальные процессы в общий процесс обучения, поскольку для подведения итогов необходимы результаты всех остальных процессов.

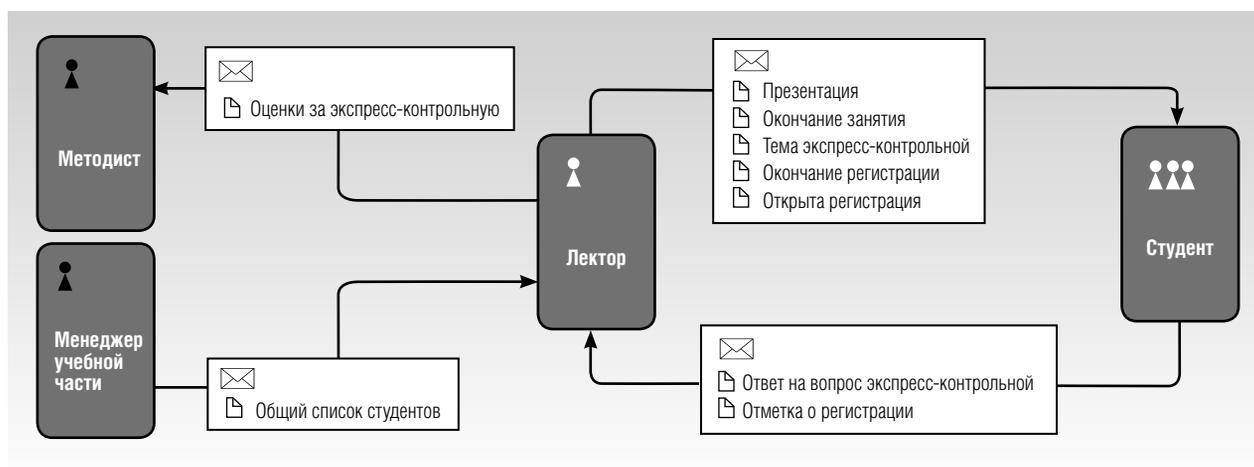


Рис. 1. Диаграмма взаимодействий процесса «Теоретическое занятие (лекция)»



страцию студентов из списка. На занятии могут регистрироваться все студенты из списка, в том числе с удаленным доступом. Участие в занятии принимают все зарегистрированные студенты, которые получают доступ к необходимым материалам, прослушивают доклад, взаимодействуют с участниками, выполняют внезапные экспресс-контрольные и др. Выполнение экземпляра процесса для лектора продолжается, пока не проверены отчеты студентов за выполнение экспресс-контрольных работ и не отправлены оценки.

Диаграмма состояний студента (рис. 3) начинается с получения сообщения о начале регистрации на лекции. Сообщение может быть получено на устройство, имеющее доступ к серверу (т.е. выход в Интернет или Интранет). Студент открывает список, регистрируется на лекции и получает необходимую информацию. Если лектор назначит контрольную работу, то студент выполняет ее и отправляет ответ лектору.

По истечении определенного времени студент получает доступ ко всем материалам лекции. Участие студента в экземпляре процесса заканчивается по окончании аудиторного занятия.

Элементы моделей (рис. 1, 2, 3) обладают свойствами, которые позволяют построить модели процесса и выполнить проектирование приложения (обычно это делает программист). Рассмотрим, как приведенные выше требования к процессу обучения реализуются в субъектно-ориентированной инструментальной среде Metasonic S-BPM Suite:

1. Быстрая адаптация процесса к изменениям обеспечивается тем, что изменения выполняют участники процесса. При этом:

- ◆ изменение диаграммы взаимодействия инициатор согласовывает со всеми участниками процесса (в режиме валидации);

- ◆ изменение диаграммы состояний не требует согласования с другими участниками процесса, кроме случая, когда эти изменения влияют на диаграмму взаимодействия;

- ◆ в особых случаях (изменение бизнес-объектов и т.п.), требующих знания языка Java, возможно привлечение программиста.

Измененные модели загружают на сервер, где хранят как версии процесса. Готовность к запуску новой версии процесса составляет от нескольких минут до нескольких часов. Инициатор запуска нового экземпляра процесса выбирает на сервере нужную версию и запускает ее на исполнение. Например, «Теоретическое занятие (лекция)» запускает «Лек-

тор», «Групповое практическое занятие (семинар)» запускает «Преподаватель», «Внеаудиторную (домашнюю) работу» запускает «Студент». Версии процесса могут быть также связаны с различными методами обучения.

2. Постоянное взаимодействие участников процесса, децентрализация полномочий, возможности для рефлексии и использования субъектности участников процесса являются неотъемлемыми свойствами субъектно-ориентированного подхода.

3. Интеграция разработанного приложения workflow с документами и информационными системами, настройка показателей эффективности процесса обеспечена возможностями инструментальной среды Metasonic S-BPM Suite и выполняется при разработке модели.

В качестве примера был рассмотрен процесс обучения конкретной дисциплине. Для управления процессом обучения другим дисциплинам необходимо разработать соответствующие модели взаимодействий и состояний с учетом уникальных особенностей дисциплины. Так же, как и для примера, необходимо разработать версии процесса с учетом особенностей исполнения отдельных экземпляров процесса. Изменения процесса могут быть выполнены самими участниками. Для начала занятия инициатор выбирает нужную дисциплину и нужную версию (или создает новую) и запускает соответствующий экземпляр на исполнение. Порядок взаимодействия (интерфейсы, последовательности) процессов согласован с расписанием и учебными планами. Таким образом, может быть выполнено моделирование и управление всеми дисциплинами, которые относятся к обобщенному процессу обучения (рис. 4).



Рис. 4. Метамоделю обобщенного процесса обучения

### Заключение

Современные требования к управлению процессом обучения делают практически невозможным использование традиционных подходов. Реализации требований возможна при использовании субъектно-ориентированного подхода и инструментальной системы Metasonic S-BPM Suite. Простота нотаций позволяет участникам процесса обучения самостоятельно вносить изменения в экземпляры процесса, создавая таким образом базу процедурных знаний. Все созданные варианты хранятся на сервере (или в облачном хранилище) и запускаются как сервисы.

Предложенная структура процесса обучения состоит из метамодели процесса и множества вариантов детальных моделей. Варианты отражают особенности дисциплины, преподавателя, аудитории, любые изменения в процессе. Использование субъектности участников процесса является основой для совершенствования процесса.

Субъектно-ориентированный подход к обучению способствует трансформации вуза от жесткой, функциональной структуры прошлого века к гибкому, сетевому рефлексивному предприятию XXI века, основанному на максимальном использовании творческих возможностей и потенциала всех участников процесса обучения. ■

### Литература

1. Лепский В.Е. Стратегичность предприятия XXI века (субъектно-ориентированный подход) // Экономические стратегии. 2006. №3. С. 110-118.
2. Рубинштейн С.Л. Человек и мир. СПб.: Питер, 2012. 224 с.
3. Чеботарев В.Г., Громов А.И. Роль субъектности в бизнес-процессах // Бизнес-информатика. 2013. №1 (23). С. 3-9.
4. Гейсельхарт Х. Обучающееся предприятие в XXI веке / Пер. с нем. Калуга: Духовное наследие, 2004. 264 с.
5. Подласый И.П. Педагогика. Новый курс: Учебник для студ. пед. вузов: В 2 кн. Кн. 1: Общие основы. Процесс обучения. М.: Гуманит. изд. центр ВЛАДОС, 1999. 576 с.
6. Давидович Н., Демьянов Е.И., Лобова Е.В., Прямикова Е.В. Стратегии вузов в обеспечении качества высшего образования // Педагогическое образование в России. 2014. №4. С. 7-16.
7. Gartner 2014. Cool Vendor [Электронный ресурс]: <http://www.metasonic.de/en/coolvendor> (дата обращения 30.09.2014).
8. Лепский В.Е., Зорина Г.И. Рефлексивное предприятие XXI века // Рефлексивные процессы и управление. 2005. Т. 5, №. 2. С. 21-40.

# AUTOMATION OF EDUCATION PROCESS

## **Valeriy CHEBOTAREV**

Associate Professor, Department of Modeling and Business Process Optimization,  
National Research University Higher School of Economics

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: vchebotarev@hse.ru

## **Alexander GROMOV**

Professor, Department of Modeling and Business Process Optimization,  
National Research University Higher School of Economics

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: agromov@hse.ru

The article is based on the results obtained by the participants of the research seminar in the faculty of business Informatics of the Higher School of Economics. Its objective is to study possibilities of learning management as a process and to apply modern information technologies and tools. The learning process is defined as joint activity of participants in organization of discipline study and obtaining of subjectively or objectively new subject and procedural knowledge. The coursework in the discipline «Modeling and optimization of business processes» has been chosen for our research. We have analyzed specifics of learning process, defined requirements of process management and made modeling and assessment of automated process management. The analysis is based on information of real studies in the Faculty of Business Informatics. We have identified the following requirements to learning process: fast adaptation to changes, flexible management using various teaching methods, constant interaction with education participants, creative and reflexive abilities using participants' personality. EEPC (ARIS) and interaction and behavior diagram (Metasonic) were used as two different notations. The first notation refers to classic modeling methodology (ARIS) and the second one – to subject-oriented methodology of process management. It was shown that subject-oriented approach meets all requirements of education process that is almost impossible with classical approach. Principles of dividing learning process into smaller processes, process models (interaction diagrams and behavior diagrams of subjects) are developed. Possibility to change process management by learning subjects (teachers, students) without programming, including the automated generation of workflow application was presented. Results of research can be useful for any organization thinking about transition from traditional rigid education structure to a more advanced reflexive learning environment with network communications.

**Key words:** learning, process modeling, process automation, subject-oriented business process management, S-BPM, flexible process, reflexivity.

## References

1. Lepskii V. (2006) Strategichnost predpriyij XXI veka [Strategic enterprise of the 21st century]. *Economic Strategies*, no. 3, pp. 110-118. (in Russian)
2. Rubinstein S.L. (2012) *Chelovek i mir* [Man and the world]. Saint Petersburg: Piter. (in Russian)
3. Chebotarev V.G., Gromov A.I. (2013) Rol' subjektnosti v biznes processe [The role of the subjectivity in business processes]. *Business Informatics*, no. 1 (23), pp. 3-9. (in Russian)
4. Geiselhart H. (2004) *Obuchajcsheesja predpriyie v XXI veke* [Learning enterprise in the twenty-first century]. Kaluga: Spiritual Legacy. (in Russian)
5. Podlasy I.P. (1999) *Pedagogika. Novi kurs* [Pedagogy. New course]. Moscow: VLADOS. (in Russian)
6. Davidovich N., Demyanov E.I., Loboza E.V., Priamnikova E.V. (2014) Strategii vuzov v obespechenii kachestva vishego obrazovanij [Strategies of universities in quality assurance of higher education]. *Pedagogical Education in Russia*, no. 4, pp. 7-16.
7. Gartner (2014) Cool Vendor. Available at: <http://www.metasonic.de/en/coolvendor> (accessed 30.09.2014).
8. Lepskii V.E., Zorina G.I. (2005) Refleksivnoe predpriyie XXI veka [Reflexive enterprise of the 21st century]. *Reflexive Processes and Control*, vol. 5, no. 2, pp. 21-40.

# FLEXIBLE ORGANIZATIONAL FORMS DESIGN BASED ON A TRANSACTION APPROACH

**Pavel MALYZHENKOV**

*Professor, Department of Information Systems and Technologies,  
Faculty of Business Informatics and Applied Mathematics,  
National Research University Higher School of Economics*

*Address: 25/12, Bolshaya Pecherskaya str., Nizhniy Novgorod,  
603155, Russian Federation*

*E-mail: pmalyzhenkov@hse.ru*

**Tatiana BABKINA**

*Senior Lecturer, Department of Applied Mathematics and Informatics,  
Faculty of Business Informatics and Applied Mathematics,  
National Research University Higher School of Economics*

*Address: 25/12, Bolshaya Pecherskaya str., Nizhniy Novgorod,  
603155, Russian Federation*

*E-mail: tbabkina@hse.ru*

**Aleksey SERGEEV**

*Post-graduate Student, Department of Information Systems and Technologies,  
Faculty of Business Informatics and Applied Mathematics,  
National Research University Higher School of Economics.*

*Address: 25/12, Bolshaya Pecherskaya str., Nizhniy Novgorod,  
603155, Russian Federation*

*E-mail: aisergeev07bi@gmail.com*

*Modern international economic environment is exposed to profound transformations of business operating conditions due to consequences of the financial crisis. Currently the organizational flexibility becomes the most important characteristic of enterprises. In its turn it presumes the adoption of such organizational structures where business relationships and aligned IT infrastructure are recognized as a specific type of the resource that a company can use to achieve competitive advantage. This research analyzes various issues of flexible organization and enterprise models which influence functionality and architecture constraints of enterprise information systems. For the analysis the authors have applied a transactions mechanism concept and specific design methodology. This paper offers an insight into key properties of four flexible organizational forms in tight connection with Enterprise Ontology formal modeling approach and DEMO, which follow the language-action perspective.*

**Key words:** organizational forms, collaborative strategies, transaction cost theory, networks, spin-offs, clusters, outsourcing, DEMO methodology, enterprise information systems.

## 1. Introduction

In the presence of the world economic crisis one of the problems frequently faced by enterprises is cost reduction and, hence, the need in optimized organizational structures to foster businesses flexibility. By flexibility we understand the capability of an enterprise to adapt rapidly their organizational structures and underlying IT infrastructure to external or internal environmental changes. Consequently, the organizational structure itself must have adequate adaptation characteristics.

As an example of such changes the customer relationship management may be considered. Customers want to increase their influence on products they purchase. Therefore organizations need to offer products and services of increasing complexity with ever shorter times-to-market. This need stimulates research in new analytical instruments and new forms of CRM systems, which facilitate dynamic restructuring of enterprise processes and even their entire organizational forms.

The lack of IS knowledge and expertise has affected especially SMEs. For example, many have not engaged in or been slow to adopt advanced IT-practices and architectural principles of information systems. A low level of organizational readiness has been suggested as the most important reason for this. Organizational readiness includes the level of knowledge about IT by managers, as well as availability of a relevant technology to develop, for example, an e-commerce website. However, there is evidence that some SMEs have been able to develop an internal level of IS expertise, particularly by gaining IS project experience over the years and by employing internal IS experts. Thus it is no longer appropriate to assume that all SMEs have low levels of internal IS expertise even if it remains the key factor for their success.

This paper amplifies approaches to select adequate organization forms and offers new principles of combining the transaction costs theory and advanced methods of business modeling based on the language action perspective. In our research we refer to Design & Engineering Methodology for Organizations – DEMO [7]. We have demonstrated that the transaction analysis represents the basis for decision support in applying of different organizational modalities of economic activity. This paper is structured in the following way: after the introduction Chapter 2 analyzes the very basis of organizational structure choice – the transactional costs. Chapter 3 provides the description of the mostly diffused flexible organizational forms and introduces the DEMO methodology. Chapter 4 outlines new results linked to

the choice of problem areas of every organizational structure, the DEMO can be applied to.

## 2. Transactional approach and its influence on organizational structures

Different authors [3, 18, 26, 27] recognize business relationships as a specific type of the resource that a company can use in the realization of relational strategies. The opportunity to mobilize others as «partners» has increasingly become an emergent issue in the strategic management literature. From a resource-based perspective the importance of business relationships is emphasized by the idea that firm's critical resources may span the boundaries of firms itself and be embedded in inter-firm resources and routines. A firm uses resources both within the firm (firm-specific resources) and at other organizations (firm-addressable resources).

Within economics there have been several approaches to the study of networks and alliances, mainly including transaction cost economics, strategic management and institutional perspective. In transaction cost economics literature various roles have been identified for management accounting in inter-firm settings that relate to specific accounting techniques and different uses of accounting information. From the economic point of view, according to the approach of transaction cost theory of Williamson [25, 27], the two fundamental ways of organizing economic activities are market and hierarchy.

The form of production coordination according to the market model is achieved when processes are broken down into individual steps carried out by independent firms that interact with each other through exchange transactions. Price formation, through the dynamics of supply and demand, ensures coordination between the parties. However, markets often do not work perfectly. Asymmetric information, uncertainty, high asset specificity and exchanged risk of opportunistic behavior are factors that determine an increase in transaction costs and force businesses to make use of hierarchy in the marketplace as a mode of organization of economic activity.

In studies influenced by the institutional theory, collaborations and networks encompass a broad range of inter-organizational relationships. Some authors [1, 21] argued that institutions supply rules and resources upon which collaboration is built. Thus, to fully understand and explore dynamics of collaboration and networks, it is crucial to examine the institutionalized patterns of rules and routines, emphasizing the objective and the

external aspects of the institutional environment.

The transaction analysis may lead to the adoption of quite different organizational solutions like networks, clusters, spin-offs or outsourcing relations. All these types are fully referred to flexible structures because they can rapidly redesign their organizational aspect as a response to external environment changes. Consequently, the information system must have similar adaptation characteristics providing high grade of flexibility. In addition, organizational restructuring has a large social aspect, therefore exploited communication paradigms, patterns and policies should be presented for decision makers explicitly during that process [23].

### 3. Flexible organizational forms

In the modern environment logistics and production flows become faster and more complex. They require more flexible management of space, time and resources in order to achieve desired levels of efficiency. From this point of view such business forms as clusters, networks, outsourcing relations and spin-offs offer different advantages. First of all, the strong socio-cultural link to a limited area (especially in the case of spin-offs and clusters) promotes rapid circulation of ideas and easy interaction between individuals who share a certain «cultural zone». It is based not only on technical and production skills sharing conveyed also through specific channels of training, but it also includes a high entrepreneurial culture and better identification of values and mutual interests of partners.

The second growth factor is the existence of a systemic approach in inter-business relations, that all the mentioned forms present, according to the logic of flexible specialization. The split nature of organizational structures often comes not from specific design patterns guided by a chief manager or head enterprises, but as a spontaneous response to the competitive environment. Thus it enables to replace a company with others, which are able to perform the same activities along the production process. At the same time, there is a remarkable stability of relationships, often based on a relationship of mutual trust, which can facilitate the search for coordination forms that can increase the overall efficiency of a business pattern.

#### 3.1 Business networks

The concept of network emerging from the collaboration strategies framework is consistent especially within the management literature focusing on «strategic net-

work». It is considered as an intentional, long-term alliance enabling different companies to acquire or defend competitive advantage against competitors outside the network. A network can be defined as «a set of business entities, legally independent, reciprocally committed to implement a deliberate and finalized cooperation strategy, leveraging the technical and economical complementarities in achieving joint economic objectives, which indirectly benefit the individual businesses» [3].

These roles include the use of financial and non-financial information in the «make or buy or ally» decisions, in the selection of a potential partner, during the management of cooperation and in the monitoring and evaluation of collaboration activities. These studies often examine inter-firm accounting in conjunction with the issue of motivation and incentives, underscoring the importance of studying accounting in a broader control context.

The subject of the strategic analysis and management accounting of networks has come to the attention of academics and practitioners over the last decades. Emergence of the global knowledge economy has triggered innovative systems and operation models, having a deep impact on business environment and organizational interdependencies. To define the current business landscape, some authors [14] suggest the metaphor of the «rainforest», indicating that one of its basic feature is the intricate interdependency among companies that requires an alternative framework to be studied.

According to this framework, strategic networks are the result of an engineered process, where the working arrangements are specified in written contracts and fulfilled in a formal organization. Besides, the goals are planned and specified in a predetermined time horizon, the cooperation is based on «network capital» rooted in business and economic rationality and focused on investing in relationships as a means to increase business performance.

The strategic management literature recognizes business relationships as another type of the resource that a company can use to achieve competitive advantage. The opportunity to mobilize others as «partners» and, hence, apply modern business planning instruments, including transaction-based ones [2, 7], has increasingly become an emergent issue in the strategic management literature based especially on the engineering approach.

Unfortunately, modern strategic context of selected agreements is mainly characterized by an «emergent» rather than «engineered» process with a negative impact on the quality of business planning that tends to be

less structured when there is no reference to an inspiring model of business interaction. The elements of business planning framework even the mandatory ones («strategic objectives» and «network action plans») present on average a poor level of elaboration, containing general intentions and planned actions directed mainly on internal processes and resources, without a clear vision of desired strategic positioning.

### 3.2 Spin-offs

The characteristics of the national innovation system of many European countries explain the serious impact of the economic crisis on innovation. Policy responses were concerned with supporting innovation systems and developing innovation capacity, such as improving infrastructure, public investments in R&D and innovation, investment in education and training at all levels, as well as demand oriented innovation policies, including public procurement, financial support to SMEs, venture capital and, an important factor, policies aimed to enterprises agglomerations development. They are seen as a part of the national strategy for coping with the effect of the financial crisis in many countries, partly because the industries involved in such programs represent industries oriented towards global markets that were most affected by the crisis [17].

The phenomenon of spin-off is characterized by the expressed willing of the employees or businessmen to create a new business distinct from the original one while they are still within this latter. It is namely this intentionality that distinguishes the spin-off from a normal formation of company by former employees. In the latter case, the output of the employees by the company is not spontaneous but it is determined by contingent reasons, related to dismissal, misunderstandings with the managerial staff or dissatisfaction with the carried out tasks. The start-up of new business venture is next to unemployment and does not imply involvement of the initial occupation. Correspondingly the growth ambitions appear less intense and motivated.

Universities and other research institutions have always given more emphasis to technology transfer mechanisms to promote cooperation between university research and industry. These alliances are very different in terms of methods and purposes, have often proved a success both for the industry that gains in competitiveness and technological advancement, and for the university that has the ability to use rich intellectual property at its disposal to finance its research and to train their students by making them more competitive and prepared for the

industrial world.

The international experience shows that namely the split off of small innovative enterprises from the big industry represents the element that links together research and business environment. They can assume the risk of transforming of business idea into industrial prototypes realization, without which it is impossible to evaluate how perspective a research idea will be in the market and if it's worth commercial realization. This mechanism is implemented by means of spin-offs. The spin-offs are majorly diffused in the USA where all the information is gathered by especially created Association of University Technology Managers.

The innovation risk is the reason due to which many large companies don't realize broad-scale investments: they need at least some guarantee of success. So, from the practical point of view the application of research to the industrial process became the niche of small innovative firms. Such procedure became possible also in Russia after the Federal Law No. 127 «About Science and State Scientific and Technical Policy» adoption. The main advantages of this Law are the following:

- ◆ unemployment reduction;
- ◆ the possibility for the universities to develop the own innovative technologies;
- ◆ enhancing of the state funds directed towards innovation development effectiveness;
- ◆ practical realization of ideas in socio-economic sphere.

### 3.3 Clusters

Cluster (industrial district) constitutes a territory with high concentration of small and medium-sized enterprises with high production specialization, generally characterized by strong interdependence of their production cycles and strongly integrated with the local socio-economic environment. The competitiveness of industrial districts derives from the particularity of productive organization in the form of flexible specialization for which the production cycle is divided into different phases and each firm is specialized in execution of a particular stage of production which ensures lower costs, flexibility and innovation [22].

Actually, clusters' formation and development became one of the priorities also at European level. Smart, sustainable and inclusive growth is key objectives indicated by the Europe 2020 Strategy paper [9]. Increased economic interdependencies and global impacts of the financial crisis demand sophisticated handling and

planning at the political level in various areas of importance. Through their value networks and proven channels between businesses, research and academic reality, clusters provide efficient catalysts for innovation policy interventions. They are able to transform policy interventions into value creation and multiply public spending by private investments.

In this context, the cluster is a valid alternative to organizational hierarchy or market, in terms of transaction costs, that can achieve the benefits of both of them. On the one hand, it allows to reach high production volumes that allow to benefit from economies of scale (a typical effect of the «hierarchy»), and on the other hand, by preserving small business district, it continues to benefit from the advantages of flexibility of its market form. The conclusion about clusters' creation and location and, more important, the aspect of interdependency of clusters participants can be made on the base of attentive transactional analysis by means of modern instruments and information systems [7].

From the business perspective it is also evident that the presence of successful clusters and clusters in crisis does not depend on the geographical location of the sectors they belong, but by realized strategic decisions. In particular, among the causes of the crisis of some districts we can see [22]:

- ◆ the inability to control the markets;
- ◆ low levels of investment in product and process innovation;
- ◆ difficulties related to generation exchange;
- ◆ the shortage of specialized skills;
- ◆ lack of cooperation between companies.

The presence of these factors eliminates the benefits of agglomeration determining in local businesses of the cluster worse performances than those of firms that operate autonomously in the same sectors. On the contrary, in case of successful clusters, enterprises adopt strategies aimed to:

- ◆ promote cooperation with other companies;
- ◆ design and implement product and process innovations by investing in research and development;
- ◆ realize joint efforts in marketing and staff training;
- ◆ strengthen the company's equity and to increase its size.

In summary, the business perspective understanding of the cluster evolutionary dynamics depends on the analysis of the behavior of individual units, which represent the real part of the explanatory capacity of growth of the entire aggregate cluster.

### 3.4 Outsourcing

Nowadays outsourcing belongs to the most actively used management organizational practices. However, implementing and sustaining change is one of the more difficult aspects of organizational leadership. It is challenging enough to align multiple stakeholders within the same company who have different priorities and goals. Outsourcing introduces additional complexities – stakeholders don't belong to the same company or organization, these can include different nationalities and cultures, and may be separated by distance, time zones, languages and commercial interests.

A possible solution for making slender the organizational structure of enterprise is the use of outsourcing of certain activities such as: services, asset management services, management information system, which can enable economies of scale and specialization, and especially lower need for employees, as that labor costs represent the largest portion of business operating costs. At the same time greater flexibility and greater propensity for technological innovation can contribute to increased productivity of labor.

Since the competition has become more complex, companies must focus on core business, that is, on those activities, where they possess specific skills, know-how and accumulated experience. Hence, it's necessary to find ideal partners at external companies to manage business areas of less strategic importance and lower value-added. Outsourcing emerges, therefore, as an organizational strategy to enhance the competitive strength of companies.

There exists a variety of outsourcing definitions. Still, all of them converge in the following: outsourcing is understood as an operating mode/strategy aimed to concentrate enterprise resources on the «core business», namely typical enterprise activities which guarantee its sustainable competitive advantage, externalizing all others.

We wish to study the phenomenon of outsourcing from the enterprise engineering point of view and here the DEMO (Design & Engineering Methodology for Organizations) represents a valid support. It is the methodology for design, engineering, and implementation of organizations and networks of organizations. Entering into commitments and complying with them is an operational principle for each organization. These commitments are established in communication between social individuals, i.e. human beings [7].

The same Prof. Dietz has incorporated the LAP-vision into the DEMO-methodology that aims to de-

liver an Enterprise Ontology, which is a constructional model of an enterprise that is fully independent on its implementation; it is dependent only on a product (structure) delivered by an enterprise. Dietz elaborates mainly how to derive this ontology from descriptions about already implemented organizations and procedures, but does not claim to cover the «brand new creation» of an enterprise from market and customer demands, although he briefly shows how to derive an ontology from a product structure. Actually drafting an organization structure based on an enterprise ontology does not enter between its scope; DEMO itself does not contain criteria or rules for implementing an organization, apart from stating a «one-actor-one-person» ideal. Furthermore, he claims an Enterprise Ontology to be a stable starting point for defining information systems

In DEMO the basic pattern of a business transaction is composed of the following three phases: the actagenic phase, when a client requests a fact from the supplier agent; the action execution that generates the required fact, and the factagenic phase, which leads the client to accept the results reported.

Basic transactions can be composed to account for complex transactions. The DEMO methodology provides an analyst with an insight of business processes of an organization, as well as agents involved. The analysis of models built on the DEMO methodology allows a company to gain detailed understanding of processes of governance and cooperation and serves as a basis for business reengineering and information infrastructure development, consistent with its business requirements.

From the enterprise engineering point of view (and DEMO, in particular) all business processes can be divided into three large parts. There are coordination and production activities that pertain to the realization of new, original material or immaterial production facts. Such activities relate to the primary function of an enterprise, its business, and have to do with the very essence of the enterprise 'being'. For that reason, they are defined as ontological.

The important second type of activities, subordinated to the first one, is that of activities concerning collecting and providing information for carrying out the coordination and production activities at the ontological level. These ones are defined as infological. Finally, the third type regards the activities that support the infological level, and concerns the form of information (data). These so-called datalogical (or

documental) level has to do with the transmission, transformation, or storage (not necessarily electronically) of data [7].

So, operatively and in the first approximation the outsourcing means the maintenance of all ontological processes/transactions within an organization and externalization of all others. From these positions the following definition can be traced:

**Definition:** outsourcing is an enterprise strategy focused on the internal management of ontological processes and complete or partial externalization of infological or datalogical ones.

However, implementing and sustaining change is one of the more difficult aspects of organizational leadership. If clients were able to achieve its cost, capability and capacity goals by itself, then outsourcing would not be required. By definition, outsourcing requires change. The initial switch from internal to external service delivery is the first and most obvious change.

The enterprise engineering approach can give also a mechanism to analyze other mentioned organizational forms. To our mind, the results of modeling can be used for selection of the most appropriate organizational form according to the *table 1*.

Table 1.

Synthesis of possible transaction-based methods application

Organizational form	Application object	Transaction costs analysis application
Clusters	Small and medium enterprises	Complementarity relations, individuation of a business leader
Outsourcing	Large companies	Business process analysis, selection of processes to externalize, individuation of typical (ontological) business processes
Networks	Small and medium enterprises	Selection of partners and projects to develop in cooperation
Spin-offs	Large, medium and small companies, universities	Individuation of projects and competences to carry out

Applicability of the proposed method was demonstrated in the case of a real car company, which faces business challenges and needs re-engineering. During development of the method the original notation of DEMO was modified in order to include developed Actor-Function Role Table, Transaction Costs Table and Restructuring Cost Table and was enhanced by modifying Actor Trans-

action Diagram (or Organization Construction Diagram). It helps to graphically represent all the changes and costs associated with enterprise restructuring.

The analogous technique can be applied to all organizational solutions described earlier. Future research may focus on particularities of transaction techniques according to application object and special organizational form and the generalization of proposed method onto all organizational solutions presented in Table 1, which represent the effective way of costs reducing and enhancing enterprise's reaction to market changes. Very often this approach application leads to small-medium enterprises creation.

Traditionally, SMEs have several advantages over a large company due to their size and flexibility in adapting to change. It has also been found that market and learning-oriented SMEs, facing strong competition, tend to be more innovative and resilient. Competition and sustainability for SMEs involve factors such as changing market trends, changing technologies and emerging new management and organizational techniques. With advances in information technology (IT)/information systems (IS), entrepreneurs are interested in developing a virtual enterprise with suitable strategic alliances that are based on core competencies. Also, the majority of SMEs are keen to adopt advanced ITs such as electronic data interchange (EDI), enterprise resource planning (ERP) and e-commerce with the objective of improving their own supply chain efficiency first and then the supply chain of their partners.

#### 4. Results and discussion

To our opinion, such innovative cross-disciplinary connection facilitates unbiased and comprehensive analysis of social interactions within current or planned organizational forms of the enterprise, as well as it offers practical advices for particular transitions or transformations of the enterprise towards a more robust and effective organization. For each of the flexible organizational forms, which have been analyzed in this article, forthcoming analysis of DEMO models provides decision makers with particular means of organizational transformations. For example, in order to smoothly transit to such forms, as spin-offs, networks or outsourcing application an analyst may use DEMO models to design the best strategy of enterprise splitting.

Obviously, such strategy deals with information systems management at small and medium enterprises, and a manager of a small business will be limited in

what activities can be initiated. Hence, he will emphasize allocating scarce resources in the near term. This may be interpreted in different ways for various stakeholders including consultants, vendors and government. Consultants must realize that the near term focus of the small business manager may not be in the best interest of the manager. The multi-staged decision making approach of the small business manager should be placed within a long-term plan which will benefit the small business.

While applying such models to the clusters analysis they help to produce an affordable plan of enterprises' merging or may provide guidelines for even more complex transformations like transformations to a holding or an industrial/financial group formation.

From such positions the use of the DEMO methodology for enterprise structure modeling and creation of flexible structures is quite advantageous. It is easily reproducible, and it can be applied regardless business segment of the enterprise. Besides, from this point of view different graphical tools like Xemod, for example, become of fundamental importance.

This paper has demonstrated how DEMO transactions facilitate comprehensive analysis of enterprises' weak points and, hence, the proper selection of suitable organizational form, according to problems faced by an enterprise. In comparison with other prevalently qualitative approaches like Delphi method, panels or expert evaluation which main weakness point is its subjectivity [10, 13, 28] our proposal relies on quantitative metrics to estimate enterprise restructuring and future operation costs, that leads to better understanding by enterprise stakeholders and more accurate and objective planning of changes.

From the economic point of view the described solution gives an opportunity of the most efficient costs control and hence the prospective of financial management procedures improvement. Further research could regard an attempt to investigate also enterprise revenues in order to complete the economic activity analysis. Certainly, such analysis can be efficiently supported by software tools. As a way for future development of instrumental support for the proposed method we consider to develop a design support software package that will help to automate and to simplify the method application. Moreover, it will reduce enterprise modeler's time for creating DEMO diagrams and estimations costs, and it will also give an opportunity to enterprise stakeholders to use this tool themselves in future and not to involve 3rd-party enterprise modeler. ■

References

1. Antonioli D., Mazzanti M., Pini P. (2010) Productivity, innovation strategies and industrial relations in SME. Empirical evidence for a local manufacturing system in Northern Italy. *International Review of Applied Economics*, no. 24 (4), pp. 453-482.
2. Babkin E., Sergeev A. (2013) Towards developing a model-based decision support method for enterprise restructuring. *Advances in Enterprise Engineering IV: Springer Lecture Notes in Business Information Processing (LNBIP)*, vol. 146, pp. 17-27.
3. Cardoni A. (2012) Business planning and management accounting in strategic networks: theoretical development and empirical evidence from enterprises' network «agreement». *Management Control*, no. 3, pp. 91-116.
4. Casson M.C. (2000) An entrepreneurial theory of the firm. Competence, Governance and Entrepreneurship: *Advances in Economic Strategy Research / Eds.: N.Foss, V.Mahnke*. Oxford: Oxford University Press, pp. 116-145.
5. Chapman R.B., Andrade K.R. (1998) *Insourcing after the outsourcing: MIS survival guide*, AMACOM.
6. Chepurenko A., Round J. (2012) Entrepreneurship in the informal sector during transition: causes, specific features, socio-economic effects. *Proceedings of XIII April International Academic Conference on Economic and Social Development. Moscow, 3-5 April 2012*, vol. 1. Moscow: HSE, pp. 443-451.
7. Dietz J.L.G. (2006) *Enterprise ontology: Theory and methodology*, Heidelberg, New York: Springer.
8. Dumay M, Dietz J.L.G., Mulder J.B.F. (2005) Evaluation of DEMO and the Language-Action Perspective after 10 years of experience. *Proceedings of the 10th International Working Conference: The Language Action Perspective on Communication Modeling. Kiruna, Sweden, 2005*. Eds. G.Goldkuhl, M.Lind, S.Haraldson. Linköping: Linköping University, pp. 77-106.
9. European Commission (2010) *Europe 2020 Strategy Paper*, European Commission.
10. Gabrielsson J., Morten H. (2004) Context, behavior and evolution: Challenges in research on boards and governance. *International Studies of Management and Organization*, vol. 34, no. 2, pp. 11-36.
11. Greaver M.F. (1999) *Strategic Outsourcing*, AMACOM.
12. Greenwood R., Suddaby R., Hinings C.R. (2002) Theorising change: The role of professional associations in the transformation of institutionalized fields. *Academy of Management Journal*, no. 45, pp. 58-80.
13. Gummesson E. (2000) *Qualitative Methods in Management Research*. Sage Publications.
14. Hakansson H., Ford D., Gadde L.E., Shenota I., Waluszewski A. (2009). *Business in Networks*, New York: John Wiley & Sons Ltd.
15. Hoogervorst J.A.P. (2009) *Enterprise Governance and Enterprise Engineering*, Springer.
16. Johanson M., Lundberg H. (2011) *Network Strategies for Regional Growth*, Palgrave: MacMillan.
17. Malyzhenkov P., Tikhobaev A. (2012) Special economic zones in Russia as an instrument of IT potential enhancing. *Proceedings of 11th International Conference «Perspectives in Business Informatics Research» (BIR 2012), Nizhny Novgorod, Russia, September 2012*. Nizhny Novgorod: National University Higher School of Economics, p. 136-147.
18. Meiseberg B., Ehrmann T. (2013) Tendency to network of small and medium-sized enterprises: Combining organizational economics and resource-based perspectives. *Managerial and Decision Economics*, vol. 34, no. 3-5, pp. 283-300.
19. *National Observatory of Italian Clusters* (2012), III Report.
20. Op't Land M. (2008) *Instrument for fast and effective splitting of organizations: Management summary of the PhD thesis «Applying Architecture and Ontology to the Splitting and Allying of Enterprises»*.
21. Parkhe A. (2003) Institutional environments, institutional change and international alliances. *Journal of International Management*, no. 9, pp. 305-316.
22. Porter M.E. (1998) *The Competitive Advantage of Nations*. New York: Free Press.
23. Sampson R. (2004) Organizational choice in R&D alliances: Knowledge-based and transaction cost perspectives. *Managerial and Decision Economics*, vol. 25, no. 6-7, pp. 421-436.
24. Searle J.R. (1969) *Speech Acts, an Essay in the Philosophy of Language*, Cambridge, MA: Cambridge University Press.
25. Williamson O.E. (1975) *Markets and Hierarchies. Analysis and Antitrust Implications*, New York: Free Press.
26. Williamson O.E. (2008) Outsourcing: Transaction cost economics and supply chain management. *Journal of Supply Chain Management*, vol. 44, no. 2, pp. 5-16.
27. Williamson O.E., Winter S.G. (1991) *The Nature of the Firm. Origins, Evolution and Development*, Oxford: Oxford University Press.
28. Yanamandram V., White L. (2006) Switching barriers in business-to-business services: a qualitative study. *International Journal of Service Industry Management*, vol. 17, no. 2, pp. 158-192.

# ПРОЕКТИРОВАНИЕ ГИБКИХ ОРГАНИЗАЦИОННЫХ ФОРМ НА ОСНОВЕ ТРАНЗАКЦИОННОГО ПОДХОДА

## **П.В. МАЛЫЖЕНКОВ**

кандидат экономических наук, PhD,  
профессор кафедры информационных систем и технологий,  
факультет бизнес-информатики и прикладной математики,  
Национальный исследовательский университет «Высшая школа экономики»  
Адрес: 603155, г. Нижний Новгород, ул. Б.Печерская, 25/12.  
E-mail: pmalyzhenkov@hse.ru

## **Т.С. БАБКИНА**

старший преподаватель кафедры прикладной математики и информатики,  
факультет бизнес-информатики и прикладной математики,  
Национальный исследовательский университет «Высшая школа экономики»  
Адрес: 603155, г. Нижний Новгород, ул. Б.Печерская, 25/12.  
E-mail: tbabkina@hse.ru

## **А.И. СЕРГЕЕВ**

аспирант кафедры информационных систем и технологий,  
факультет бизнес-информатики и прикладной математики,  
Национальный исследовательский университет «Высшая школа экономики»  
Адрес: 603155, г. Нижний Новгород, ул. Б.Печерская, 25/12.  
E-mail: aisergeev07bi@gmail.com

Современная экономическая среда подвержена глубоким преобразованиям условий функционирования бизнеса вследствие финансового кризиса. В настоящее время и организационная гибкость становится наиболее важной характеристикой предприятий. В свою очередь, это предполагает принятие предприятиями таких организационных структур, в которых бизнес-отношения и ИТ-инфраструктура признаются как специфический тип ресурса, который может использоваться компанией для достижения конкурентного преимущества. Настоящее исследование анализирует различные аспекты гибкой организации и моделей предприятий, которые влияют на функциональность и архитектурные ограничения информационных систем предприятия. Для анализа авторы прибегают к концепции транзакционного механизма и особой методологии проектирования. Статья предлагает рассмотреть ключевые свойства четырех гибких организационных форм в тесной связи с подходом формального моделирования Онтологии Предприятий и методологии ДЕМО, которые следуют перспективе «язык-действие».

**Ключевые слова:** организационные формы, совместные стратегии, теория транзакционных издержек, сети, спин-офф, кластеры, аутсорсинг, методология ДЕМО, корпоративные информационные системы.

## Литература

1. Antonioli D., Mazzanti M., Pini P. Productivity, innovation strategies and industrial relations in SME. Empirical evidence for a local manufacturing system in Northern Italy // International Review of Applied Economics. 2010. No. 24 (4). P. 453-482.
2. Babkin E., Sergeev A. Towards developing a model-based decision support method for enterprise restructuring // Advances in Enterprise Engineering IV: Springer Lecture Notes in Business Information Processing (LNBIP). 2013. Vol. 146. P. 17-27.
3. Cardoni A. Business planning and management accounting in strategic networks: theoretical development and empirical evidence from

- enterprises' network "agreement" // *Management Control*. 2012. No. 3. P. 91-116.
4. Casson M.C. An entrepreneurial theory of the firm // *Competence, Governance and Entrepreneurship: Advances in Economic Strategy Research* / Eds.: N.Foss, V.Mahnke. Oxford: Oxford University Press, 2000. P. 116-145.
  5. Chapman R.B., Andrade K.R. *Insourcing after the outsourcing: MIS survival guide*, AMACOM, 1998.
  6. Chepureno A., Round J. Entrepreneurship in the informal sector during transition: causes, specific features, socio-economic effects // XIII Апрельская международная научная конференция по проблемам развития экономики и общества. Москва, 3-5 апреля 2012 г. М.: ВШЭ, 2012. Кн. 1. С. 443-451.
  7. Dietz J.L.G. *Enterprise ontology: Theory and methodology*. Heidelberg, New York: Springer, 2006.
  8. Dumay M., Dietz J.L.G., Mulder J.B.F. Evaluation of DEMO and the Language-Action Perspective after 10 years of experience // *Proceedings of the 10th International Working Conference: The Language Action Perspective on Communication Modeling*. Kiruna, Sweden, 2005. Eds. G.Goldkuhl, M.Lind, S.Haraldson. Linköping: Linköping University, 2005. P. 77-106.
  9. *Europe 2020 Strategy Paper*, European Commission, 2010.
  10. Gabriellsson J., Morten H. Context, behavior and evolution: Challenges in research on boards and governance // *International Studies of Management and Organization*. 2004. Vol. 34, no. 2. P. 11-36.
  11. Greaver M.F. *Strategic Outsourcing*. AMACOM, 1999.
  12. Greenwood R., Suddaby R., Hinings C.R. Theorising change: The role of professional associations in the transformation of institutionalized fields // *Academy of Management Journal*. 2002. No. 45. P. 58-80.
  13. Gummesson E. *Qualitative Methods in Management Research*. Sage Publications, 2000.
  14. Hakansson H., Ford D., Gadde L.E., Shenota I., Waluszewski A. *Business in Networks*. New York: John Wiley & Sons Ltd., 2009.
  15. Hoogervorst J.A.P. *Enterprise Governance and Enterprise Engineering*. Springer, 2009.
  16. Johanson M., Lundberg H. *Network Strategies for Regional Growth*. Palgrave: MacMillan, 2011.
  17. Malyzhenkov P., Tikhobaev A. Special economic zones in Russia as an instrument of IT potential enhancing // *Перспективные направления исследований в области бизнес-информатики (BIR 2012): Материалы XI международной конференции, Нижний Новгород, Сентябрь 2012*. Нижний Новгород: ВШЭ, 2012. С. 136-147.
  18. Meiseberg B., Ehrmann T. Tendency to network of small and medium-sized enterprises: Combining organizational economics and resource-based perspectives // *Managerial and Decision Economics*. 2013. Vol. 34. Issue 3-5. P. 283-300.
  19. *National Observatory of Italian Clusters. III Report*. 2012.
  20. Op't Land M. Instrument for fast and effective splitting of organizations: Management summary of the PhD thesis "Applying Architecture and Ontology to the Splitting and Allying of Enterprises". 2008.
  21. Parkhe A. Institutional environments, institutional change and international alliances // *Journal of International Management*. 2003. No. 9. P. 305-316.
  22. Porter M.E. *The Competitive Advantage of Nations*. New York: Free Press, 1998.
  23. Sampson R. Organizational choice in R&D alliances: Knowledge-based and transaction cost perspectives // *Managerial and Decision Economics*. 2004. Vol. 25. Issue 6-7. P. 421-436.
  24. Searle J.R. *Speech Acts, an Essay in the Philosophy of Language*. Cambridge, MA: Cambridge University Press, 1969.
  25. Williamson O.E. *Markets and Hierarchies. Analysis and Antitrust Implications*. New York: Free Press, 1975.
  26. Williamson O.E. Outsourcing: Transaction cost economics and supply chain management // *Journal of Supply Chain Management*. 2008. Vol. 44, issue 2. P. 5-16.
  27. Williamson O.E., Winter S.G. *The Nature of the Firm. Origins, Evolution and Development*. Oxford: Oxford University Press, 1991.
  28. Yanamandram V., White L. Switching barriers in business-to-business services: a qualitative study // *International Journal of Service Industry Management*. 2006. Vol. 17, issue 2. P. 158-192.

# STANDARDIZATION IN SOFTWARE PRODUCTION AT THE CORPORATE LEVEL: RESULTS OF RESEARCH IN CIS

**Denis PASHCHENKO**

*Independent consultant in software domain  
E-mail: denpas@rambler.ru*

**Andrey BLINOV**

*Professor, Department of General Management, Faculty of Management,  
Financial University under the Government of Russian Federation;  
Corresponding Member of the Russian Academy of Natural Science  
Address: 49, Leningradskiy Prospect, Moscow, 125993, Russian Federation  
E-mail: aoblinov@mail.ru*

*This paper focuses on challenges associated with modification and enhancement of process models in software production in CIS region, usually accompanied by specific risks and organizational resistance, and aggravated by weakness of formal corporate change management structures. All findings and conclusions are based on authors' survey carried out at the end of 2013 that covered 21 managers of software companies from CIS. The study was aimed to address challenging issues of software production standardization and certification, organizational resistance and other specifics of change management at the level of an entire company. This paper highlights relevant institutional interventions to support change management at planning, staff preparation and change implementation phases. The experts have ascertained that the systemic approach to change management is needed, including formal change planning activities and establishment of a special team change management for an internal project. Also the experts have shared their practical experiences and outputs: typical challenges, change reinforcement techniques and transformation timeframes.*

*The authors have resumed their research findings by formulating the following recommendations: to use a 4-stage lifecycle change plan, to manage general and specific risks at all stages, to formalize change management and to use change implementation analysis results in future practice.*

**Key words:** changes implementation management, software production improvement, organizational resistance in software company.

## Introduction

The complexity of standardization in software development is a well-known problem of the IT industry in all over the world. In CIS (Commonwealth of Independent States) region a part of that evolution process was missed at the end of 1990s, when new and progressive ISVs (Independent Software Ven-

dor) and outsourcing companies implemented advanced process models based on CMM (Capability Maturity Model) and RUP (Rational Unified Process). There are also plenty of IT-companies in CIS countries, which have built their own process models of software production themselves, basing them on habits of management, sometimes without taking into account end-customer

expectations. On the other hand, over the last 10 years software companies - newcomers have tried to use agile and hybrid methodologies. In the authors' survey over-all experiences and opinions of 21 experts from different kinds of software companies have been grouped and identified:

- ◆ effective approaches in process improvement and change management;
- ◆ key factors of resistance and cooperation of participants in processes improvement;
- ◆ possible future scenarios of software development improvement.

Meanwhile, IT-companies from CIS (and first of all from Russia) have been playing an increasingly important role in the world market software development and have been enjoying a rapidly growing share. It means that success in production and business improvement at these companies has a strong impact on the regional economy.

The IT industry has been evolving very rapidly thanks to its technologies, automation tools, modern methodologies, educational standards and end-customer expectations. It means that production processes should be flexible and capable to accommodate to rapidly changing environment [1]. Proven approaches and practices in change implementation give additional chances for successful production, business improvement, and meeting customer requirements.

A Russian software production enterprise is an interesting field to observe, how modern approaches in change management and software development process standardization gain new highlights and specific aspects in half-isolated conditions.

### 1. Research method and process

The survey was conducted in the period since September 9 till December 18, 2013 by a 3 round-Delphi study. Twenty-one Russian speaking experts from CIS-countries took part in the survey. All experts were leading managers at their companies: from project managers with team of 15+ people to software quality directors and CEO of software companies with hundreds software engineers.

In the first round the panelists sent their opinion and answers on the list of questions split into 3 sections:

- ◆ common questions about influence of production processes standardization and company certification on the quality of software products;
- ◆ special questions about experience and best prac-

tices at the level of production of the whole company;

- ◆ prognosis and opinions concerning 10 years perspective of software development process models and instruments in CIS countries.

In the second round the panelists received principal opinion of experts' panel for each of the questions. If expert's answer differed from the principal opinion, the expert could correct his answer or just give a comment.

In the third round the panelists gave additional information and comments that helped to improve Delphi study results and objectivity.

The process of gathering experts' opinions is worth being described in more details in this article, as well as generalization of the results in the form of ranked lists and bar/pie charts.

When the responses collected were analyzed during the first round, for each question the dominant (principal) opinion was selected to become the general consensus of the panel. In the 2nd round the responses of each expert were compared with the principal opinion of the panel to provide an expert with an opportunity to change or to comment.

As a result, for every multiple-choice question a ranked list with the dominant response was received at the beginning. If expert's answer was not in the top of answers list, then in the 2nd round comment from his side was requested.

For questions with one possible embodiment of the response charts were built to demonstrate popularity of answers in percents. It helps to receive the whole panel's opinion and to further develop the methods and recommendations.

The following *table* contains the number of active experts for each of the rounds.

*Table 1.*

**Activity of experts for rounds  
of the Delphi study**

	Round 1	Round 2	Round 3
<b>Active experts</b>	21	16	21
<b>Percent of active experts</b>	100%	76%	100%

In round 2 we faced obvious decrease of expert's activity.

The following charts show different information about the experts, their experience and geographical locations.

Presented experience is usually most relevant for the same type of IT-companies [2]. Types of IT-companies were presented in Delphi Panel in the following ratio:

- ◆ 10% of the experts with experience at non IT-companies with in-house development;
- ◆ 14% of the experts with experience at software vendors (ISV);
- ◆ 29% of the experts with experience at software system integrators;
- ◆ 48% of the experts with experience at tailor-made software companies (include out-sourcing model).

CIS-region geography of the survey is presented below:

- ◆ 57% of the experts from Moscow and Sankt-Petersburg (Russia);
- ◆ 19% of the experts from other cities from Russia;
- ◆ 14% of the experts from Ukraine;
- ◆ 10% of the experts from other countries of CIS region.

The experts presented the middle age group, usually considered in the IT area as the apex of creativity and professional activity:

- 5% of the experts at age 20-29;
- 62% of the experts at age 30-39;
- 33% of the experts at age 40-49;
- 0% of the experts at age 50+.

Meanwhile, most of the experts had been working in software development area for considerable number of years, so there were no experts in the panel working in the IT sector for less than 5 years:

- ◆ 19% experts had been working in the software development area from 5 to 10 years;
- ◆ 81% experts had been working in software development area for more than 10 years.

## 2. Results

### Section 1. Overview of standardization of software production

In this section the experts answered questions relating to importance of software production process standardization and official company's certification. Also in this section detailed perspectives of process model improvement were included.

According to the principal opinion of the panel, there is a strong and visible correlation between software development process standardization and final quality of software products. Most of the experts consider that standardization is a key factor of high quality in software development. On the following diagram and all others

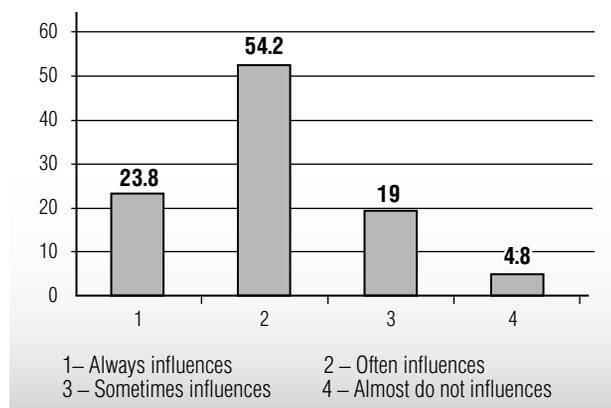


Fig. 1. How often does software development standardization influence the final quality of software product?

numbers mean percent of experts that have chosen the relevant answer (Fig. 1).

Actually it implies that one of approaches of company's management to impact the quality of software production is to establish a quality control department that should work over standardization of process model and improve quality of software products. CIS experts rely on standardization as a key method for product quality improvement with some additional remarks: time schedule and even opportunity of process standardization depends on many factors, such as:

- a. Qualification of project office and software production management;
- b. Common character of all produced software products.

The principal opinion of the experts about obligatory certification for a company is not so clear. About 25% of the experts have not seen any relation between well-known certifications or appraisal (like ISO, IC Agile or CMMI) and high quality of software products. A big share (about 43%) of the experts has considered that official certification could confirm high quality of software only sometimes. Meanwhile, some experts have remarked that even preparation for certification and first audits may temporary increase software quality. But typical opinions were quite different:

«... often company do not support requirements between certifications and do not manage the quality of software...» or «there is a lot of cases then CMMI Appraised software companies do not support their own standards in more than one reviewed project».

Experts' opinions about predominance of any kind of process models in software development in perspective of 10 years were different as well. But a trend of popularity decreasing for classic iterative software development

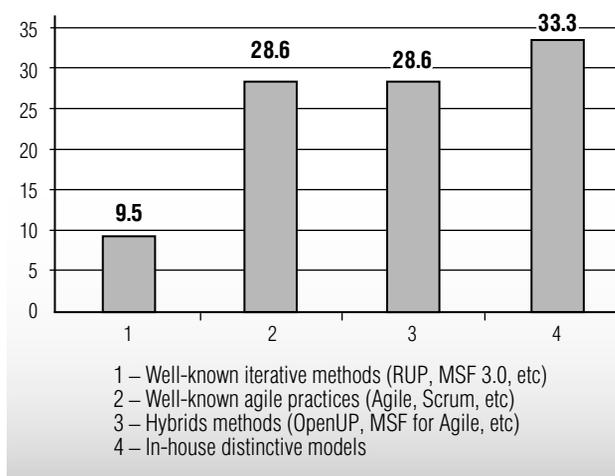


Fig. 2. Which methodologies and standards from your point of view are promising in the next 10 years for practical usage in software commercial development?

models (like RUF or MSF) appears in CIS countries, and now much more attention is paid to hybrid and agile methodologies. However, a significant share of market would be taken by companies with their own vision of software production models, including picking elements from iterative, flexible and hybrid models. Certainly, last variant means significant predominance of own distinctive expertise of project teams and specialists, rather than simply adaptation of existing standards to a specific software product or region (Fig. 2).



Fig. 3. Is it necessary to do regular process reengineering in software production?

The experts haven't come to a common opinion about needs of regular reengineering of current production process model. But the majority of experts have found that improvement of software production should have been regular even if it has not been done by drastic reengineering (Fig. 3).

In addition to above mentioned, the panel has found that for the CIS companies there have been a lot of

well-known cases, when standardization driven from the center had led companies back to decentralization in quality management (for example, in program of projects or business division). After such kind of decentralization the improvement of processes had been executing at the level of projects or directions.

It means that only dedicated organization units like SEPG (Software Engineering Process Group) or quality management direction may focus all efforts on continuous process improvement or even just on compliance with accepted production standards. Also such kind of unit may proactively audit the needs of partial or full reengineering of process model in software production.

### Section 2.

#### Changes in processes of software production at the whole company level

In this section the experts shared their opinions and experience concerning practices of changes implementation in software development processes at the level of the whole company – or at a separated division (subsidiary), focused on software development at a company. Of course, this experience related to significant changes that have impacted all stages of processes and all the project team members. For example, this kind of change could be implementation of CMMI principals in production or a new approach in usage of agile practices.

Of course, implementation of any organizational improvements starts from the planning stage and includes estimation of its expected effects [3]. The experts have come to an agreement that for an internal project of software production process improvement it is strictly obligatory to have the full and actual set of project documenta-

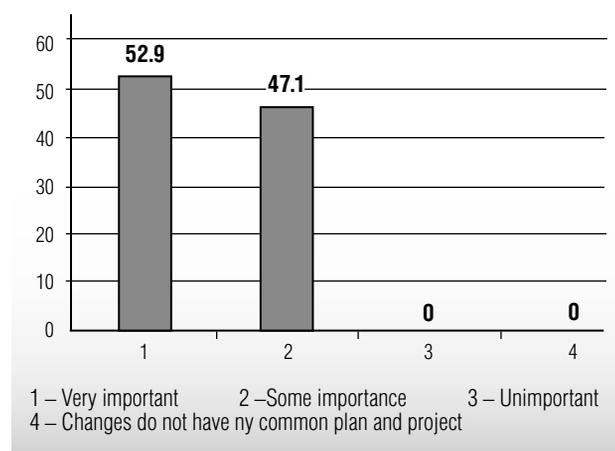


Fig. 4. How important is to have a full and actual set of project documentation in internal process improvement in software production (project plan, risk table, resource map, etc)?

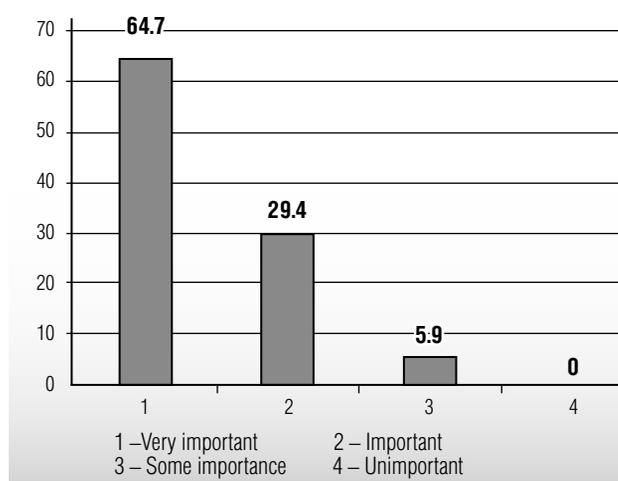


Fig. 5. How important is to have the formal phase of planning in project schedule of changes implementation in software production at the level of the whole company?

tion (like project manifest, plan, risk table, etc.) (Fig. 4).

It is also important to have a formal phase of planning in such internal project (Fig. 5).

The experts have found the following list of arrangements required to get company's staff ready for future organizational and process changes in software development (given in order of popularity, but all the actions presented in Delphi study are relevant):

- ◆ Kick-off meetings and detailed explanation for the whole staff;
- ◆ Personal meetings with line and project managers;
- ◆ Internal marketing support of future changes;
- ◆ Announcing of changes by the top management.

Thus kick-off meeting is the most common and popular practice that is used in practice of more than 90% of the experts. Strange as it may seem, the «internal marketing support» was met only in experience of 47% of the

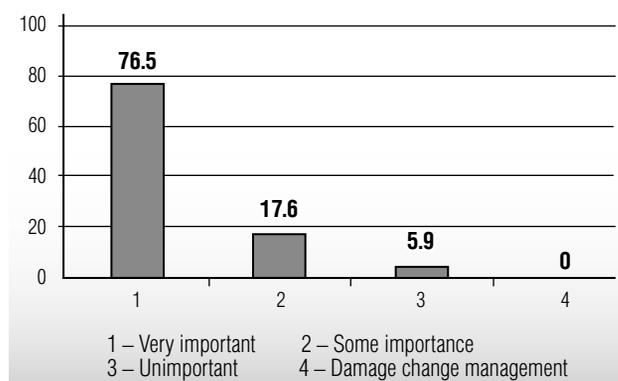


Fig. 6. How important is to form a separate team for internal project of production processes improvement (with participation of different company's managers)?

experts [4], meanwhile «support of ordinary engineers is a well-known factor of success in change management at IT-companies».

Also the panel was sure that a dedicated team for internal project of process improvement should have been allocated. This team should have been formed from managers and leading specialists for whom this team's activities were not the main job at a company, but would rather have complemented their basic functions (Fig. 6).

The experts have also identified the dominant role of company's first person (CEO) for initiating and implementing changes in production processes at the level of the whole software company: in practice of more than 90% of the experts a top-manager significantly helps to overcome internal project's crises, such as personal conflicts or lack of resources. Only 6% of the experts, however, have met a CEO, who has managed this kind of project in software companies directly.

Involvement of the top management at early stages of internal process improvement gives strong benefits and helps to overcome a lot of regular problems. At a company innovators shouldn't be afraid of high expectations or super extra pushing from the top management side, because they rather prefer to watch process of improvement from a distance and correct it only in special cases.

The panel has agreed that a major and most frequently recurring problem was the problem of formal attitude from the side of process participants. This formal attitude means change implementation without significant results and real understanding of its main goals. More than 80% of the experts have faced such kind of problem in their practice. Meanwhile, more than 50% of the experts have encountered huge resistance of IT-company staff, involved in changes of business processes.

It means that explanation and wide debates about goals and process of changes implementation should be started at early stages and should continue through the whole project. There are lots of practices and approaches supporting involvement of company's staff in Total Quality Management [5] or compliance with the standardized processes. Those of them which are relevant for a case should become regular activities in appropriate internal project plan.

The experts have also identified a problem of serious contradictions between current practices in projects at various stages and new approaches that lead to simultaneous maintenance of several different methodologies at a company; obviously it takes more efforts and frustrates

employees. At the same time «imaginary» interest of the top management declaring the high priority of changes, but at crucial moments not supporting the corresponding project, causes significant difficulties.

Generally, the panel has not defined the vector of current customers' influence on internal projects of changes implementation. A significant part of the experts (around 30%) considered that the current stakeholders, interested in a final software product, had not impacted on such kind of internal activities.

The panel has compiled an agreed list of cardinal methods of overcoming staff resistance at software companies (in order of method's popularity):

- ◆ Involvement of resisting persons into changes implementation;
- ◆ Positive motivation of the staff to adopt changes;
- ◆ Advocacy with elements of suppression.

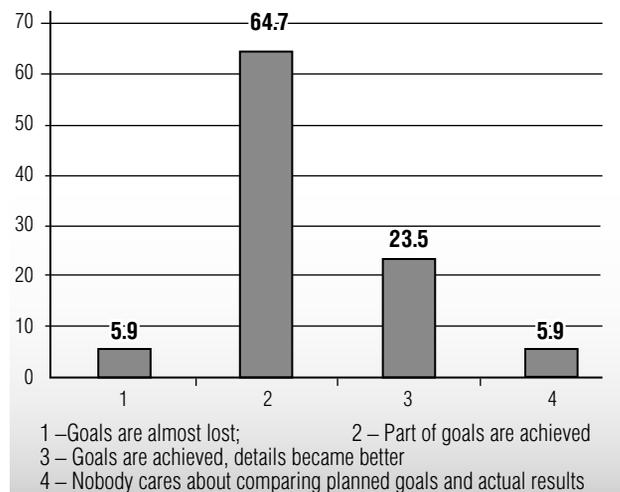


Fig. 7. How strong are planned goals of changes implementation usually modified at the end of an internal project?

Meanwhile, only 10% of the experts have admitted direct suppression as a useful method.

It's important to compare planned and actual results of changes implementation. Experience of the panel was quite positive (Fig. 7).

On practice it means, that the original planned goals may be sorted in groups and for each group of the goals a separate project stage in process improvement may be scheduled.

Also the time schedule of internal project at a dynamic IT enterprise is really an important project parameter. Experience of the panel shows, that in most cases the originally planned dates have been usually exceeded (Fig. 8).

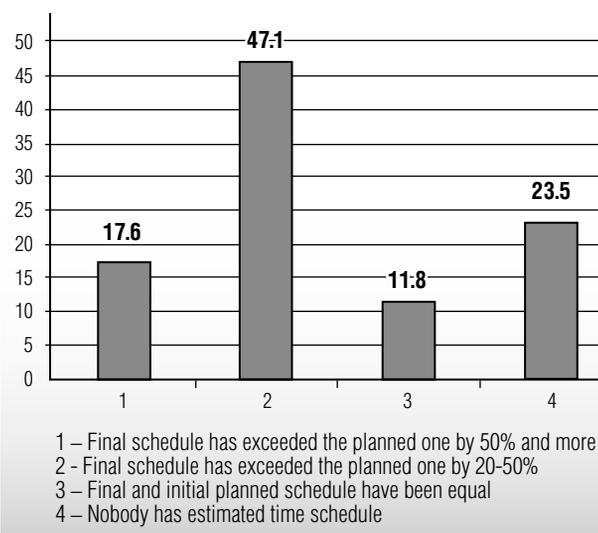


Fig. 8. How is final time of changes implementation in the production processes at the level of the whole company comparable with planned schedule of internal project?

Thus the additional risk of the lack of time is relevant for such kind of projects, and monitoring of project schedule should be regular and pro-active. At the planning stage a management team should consider possible time reserves.

The panel has agreed that it was important to arrange a formal assessment of change implementation in software production processes at the company level (Fig. 9).

It means that summarizing at the end of an internal project should be formal and should be scheduled in a project plan. Also such kind of reports may be reviewed from time to time, especially at the beginning of next stages of an internal process improvement project.

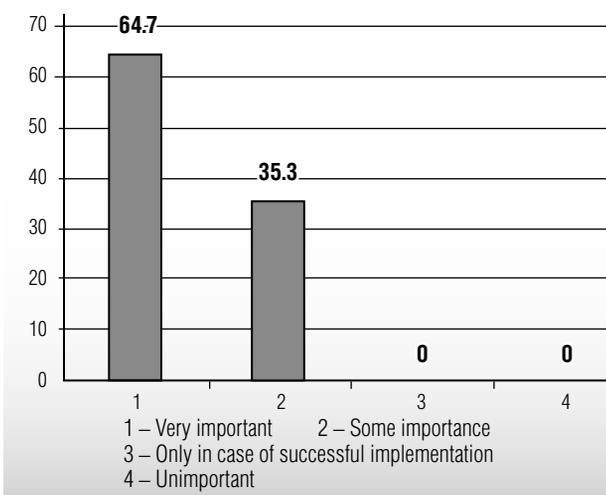


Fig. 9. How important is formal summarizing of results of change implementation in software production processes at the company level?

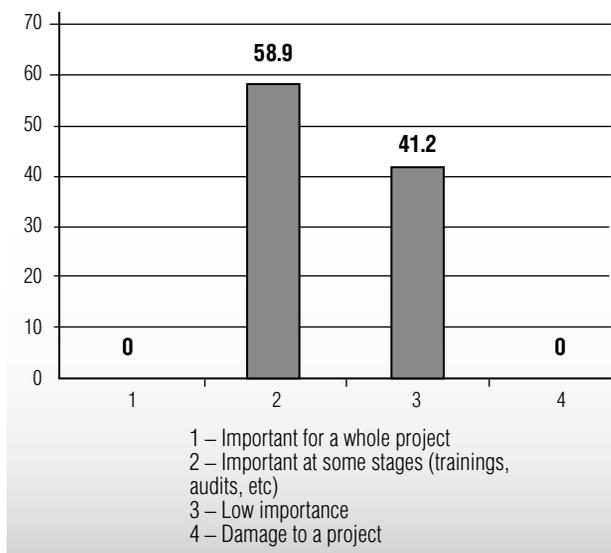


Fig. 10. How important is involvement of external consultants in internal project of change implementation in production at the level of the whole software company?

The experts have defined a set of effective steps to reinforce implemented changes in production practice of a company (in order of answer's popularity):

- ◆ Additional audit of process execution;
- ◆ Process documentation respecting corporate standards and instructions;
- ◆ Attention to process or practice in case of recurring defects and problems;
- ◆ Internal marketing support of implemented changes.

Also the experts have added some actions into the list of best practices:

- ◆ Internal trainings;
- ◆ Automation tools configured according to the new processes.

The panel has mainly identified the role of external consultants in internal projects of production process improvement at the level of the whole company as «important at some stages» (Fig. 10).

### 3. Conclusions and recommendations

The experts' panel has recommended to use the same or wider set of documentation for an internal process improvement project, than for external software / consulting projects. For such kind of project dedicated team from leading specialists and managers should be allocated, executing these roles at company in addition to their basic functions.

Based on the survey findings the authors may also recommend to pay attention to the formal stage of planning, when a manager of this kind of internal project may spend time on risks management and planning important items:

- ◆ Additional time reserves;
- ◆ Involvement of external consultants at some stages and activities (like training or audits);
- ◆ All arrangements and actions aimed to overcome typical implementation problems;
- ◆ To gain support and loyalty of top managers, who may help passing critical points of a project.

There are two well-known problems in such kind of projects that may be envisaged at the planning stage: lack of time and lack of resources. Additional time reserves could help to mitigate the first risk, and involving top managers could help to address the second one. Support of top managers (like CEO, CTO or COO) could be a strong helping factor, giving additional chance for success to a software production process improvement project. Involvement of top managers into change management at a high level may be the most valuable resource at this stage [6].

Dividing an internal project into phases at its planning phase could help to prioritize its goals and to get additional chances for successful achieving at least part of them.

At the next formal stage of an internal project – preparation of company's staff for future changes [7] – the panel has recommended to start a set of activities, first of all: kick-off meetings and detailed explaining for all the staff. Well-done preparation of employers for future changes may save a lot of time and efforts for innovators at the next stages of an internal project – a detailed study of changes and changes implementation.

Change implementation faces a lot of risks and problems at IT-companies [8]. This Delphi study has shown some of these problems, e.g. formal implementation without results and without its understanding by employees, and even organizational resistance. It requires a lot of efforts and attention during all implementation stages from an internal project team. The experts have recommended:

- ◆ Involvement of resisting persons into change implementation;
- ◆ Positive motivation to adopt changes;
- ◆ Advocacy with elements of suppression.

Based on the survey findings the authors may also recommend formalization and documentation of internal project outputs no matter on its results. Such kind

of report may be used to plan future process improvement or to implement correction activities at next stage of change implementation.

The experts have recommended the set of effective steps to assess implemented changes in production practice of a company, such as:

- ◆ Additional audit of process execution;
- ◆ Process documentation respecting the corporate standards and instructions;
- ◆ Attention to process or practice in case of recurring defects and problems;
- ◆ Internal marketing support of implemented changes.

The survey has shown the importance of process im-

provement and standardization that needs planned and balanced approach for change implementation at the level of the whole company. The panel responses, especially in consensus opinions, have demonstrated the necessity of considering all the factors of organizational resistance and analysis at each stage of a change implementation project.

The study performed has revealed that the main ideas and approaches in software production standardization have strong reflection in practice of software companies in CIS region – even respecting such CIS-specific factors as ownership structure, strong language barrier with world's software standards & experts, old-style traditions for information technologies. ■

### References

1. Lipaev V. (2008) *Ekonomika proizvodstva slozhnykh programmnykh productov* [Economical problems of production of complex software products]. Moscow, SINTEG. (in Russian)
2. Parabelum A., Zapirkin D (2011) *Razvitiye biznesa* [Business Development]. Moscow, Infobusiness. (in Russian)
3. Pomeroy-Huff M., Mullaney J., Cannon R., Sebern M. (2005) The Personal Software Process (PSP) Body of Knowledge, ver. 1.0. Special report CMU/SEI.
4. Tsipes G., Kuzmishchev A. (2014) Proekty organizacionnykh izmeneniy v krypnykh kompaniyah: metody ocenki i prinyatiya resheniy [Organizational change projects in large companies: assessment and decision-making methods]. *Project and Program Management*, no. 1, pp. 6-21.
5. Fowler P., Rifkin S. (1990) Software Engineering Process Group Guide CMU/SEI-90-TR-024. Pittsburg, Carnegie Mellon University.
6. Pashchenko D. (2013) Rol' rukovoditelya IT-kompanii v processe vnedreniya izmeneniyv proizvodstvennyye processy [The principal role of CEO of IT-company in the implementation of changes in the production processes model]. *The News of Tula State University*, vol. 4, part 1, pp. 164-179.
7. Pashchenko D. (2012) Proektirovanie organizacionnykh izmeneniy v IT-kompaniyah s uchedom faktorov protivodejstviya [Resistance to organizational changes in software development companies]. *Management and Business Administration*, no. 4, pp. 170-179.
8. Mosyagin M. (2012) *Analiz neobhodimosti i tekhnologiya organizacionnykh izmeneniy v IT-kompanii* [Analysis and technologies of organizational changes in IT-companies]. Theses of program MBA. IT Management School. Available at: <http://journal.itmane.ru/node/638> (accessed 10 June 2014).

# СТАНДАРТИЗАЦИЯ ПРОИЗВОДСТВА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА КОРПОРАТИВНОМ УРОВНЕ: РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ В СНГ

## **Д.С. ПАЩЕНКО**

кандидат технических наук, MBA, независимый консультант в области  
разработки программного обеспечения  
E-mail: denpas@rambler.ru

## **А.О. БЛИНОВ**

доктор экономических наук, профессор кафедры общего менеджмента,  
факультет менеджмента, Финансовый университет при Правительстве РФ;  
член-корреспондент Российской академии естественных наук (РАЕН)  
Адрес: 125993, г. Москва, Ленинградский проспект, д. 49  
E-mail: aoblinov@mail.ru

Данная статья посвящена проблемам внедрения изменений и улучшений в процессные модели производства программного обеспечения (ПО) в регионе СНГ, обычно сопровождаемые специфическими рисками и организационным сопротивлением при слабости формальных корпоративных структур управления изменениями. Все выводы и заключения основаны на авторском исследовании, проведенном в конце 2013 года и охватившем 21 руководителя софтверных компаний из СНГ. Исследование было направлено на решение актуальных проблем стандартизации и сертификации производства ПО, организационное сопротивление и другие особенности внедрения изменений на уровне всей компании. В статье приведены необходимые организационные меры, поддерживающие внедрение изменений на этапах планирования, подготовки коллектива и проведения самих изменений. Эксперты установили важность системного подхода к внедрению изменений, включая активности по формальному планированию изменений и созданию отдельной команды для внутреннего проекта. Также эксперты рассказали о своем практическом опыте и результатах: типичных проблемах, методах закрепления изменений, сроках преобразований. Авторы резюмировали итоги исследования, рекомендовав использовать четырехстадийный жизненный цикл изменений, управление общими и специфическими рисками на всех стадиях, формализацию управления изменениями и использование результатов анализа достигнутых изменений в будущей практике.

**Ключевые слова:** управление изменениями, улучшение производства ПО, организационное сопротивление в софтверных компаниях.

## **Литература**

1. Липаев В.В. Экономика производства сложных программных продуктов. М: СИНТЕГ, 2008.
2. Парабелум А., Зариркин Д. Развитие бизнеса. М.: Инфобизнес, 2011.
3. Pomeroy-Huff M., Mullaney J., Cannon R., Sebern M. The Personal Software Process (PSP) Body of Knowledge, ver. 1.0. Special report CMU/SEI, 2005.
4. Ципес Г.Л., Кузьмищев А.В. Проекты организационных изменений в крупных компаниях: методы оценки и принятия решений // Управление проектами и программами. 2014. №1. С. 6-21
5. Fowler P., Rifkin S. Software Engineering Process Group Guide CMU/SEI-90-TR-024. Pittsburg: Carnegie Mellon University, 1990.
6. Пашенко Д.С. Роль руководителя IT-компания в процессе внедрения изменений в производственные процессы // Известия ТулГУ. Экономические и юридические науки. Вып. 4., Ч. I. — Тула: Изд-во ТулГУ, 2013. С. 164-179.
7. Пашенко Д.С. Проектирование организационных изменений в IT-компаниях с учетом факторов противодействия // Менеджмент и бизнес-администрирование. 2012. №4. С. 170-179.
8. Мосягин М. Анализ необходимости и технология организационных изменений в IT-компаниях // Школа IT-менеджмента [Электронный ресурс]: <http://journal.itmane.ru/node/638> (дата обращения 10.06.2014).

# WORLD DEMOGRAPHIC SITUATION FROM THE PERSPECTIVE OF GLOBAL DEMOGRAPHIC BALANCE

**Tatiana BOGDANOVA**

Associate Professor, Department of Business Analytics,  
Faculty of Business Informatics, National Research University Higher School of Economics  
Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation  
E-mail: tanbog@hse.ru

**Tatiana YAKOVETS**

Corresponding member of Russian Academy of Natural Sciences, Doctoral candidate of Institute  
of social and economic problems of the population of the Russian Academy of Sciences  
Address: 32, Nahimovskij prospect, Moscow, 117218, Russian Federation  
E-mail: tzag@mail.ru

*In 1971 in his Nobel lecture Simon Kuznets noted that the population growth had been ceasing to be the main force of the economic growth over the last one or two decades. Accordingly, the authors have examined the contemporary demographic situation in the world based on information given in United Nations (UN) population prospects. This paper describes the global demographic balance method that includes 5 age cohorts of the population of 20 countries and regions of the world for the last period of 1950-2010 and UN Population Prospect to 2050. This method has been applied to analyze quantitative parameters of the demographic situation in developed, least developed and in developing countries. Developed countries, which had passed the demographic transition, will face a depopulation pattern in XXI. The age structure of depopulation trends in these countries is given. In least developed countries the population growth has been persisting but not at so high pace as it was in the second half of the 20<sup>th</sup> century. BRICS countries stand out from developing countries. To assess qualitative characteristics of countries of the world the Human Development Index (HDI) has been used. This paper outlines characteristics of this indicator given by United Nations Development Programme (UNDP). HDI values for BRICS countries are specified, and a conclusion has been drawn that qualitative growth is needed for economies of these countries. Outputs of world population simulation and projections by G.P. Gorshkov, B.M. Dolgoplov and A.A. Akayev, adjusted for the biosphere ultimate capacity, are presented. A conclusion has been formulated that projections by S.P. Kapitsa and UN experts that disregard the biosphere ultimate capacity are more realistic.*

**Key words:** UN demographic prospects, global demographic balance, age population structure, least developed and developing countries of the world, BRICS countries, Human Development Index, United Nations Development Programme, biosphere ultimate capacity.

## 1. Introduction

In 1971 Simon Kuznets said in his Nobel Prize lecture: «Six characteristics of modern economic growth have emerged in the analysis based on conventional measures of the national product and its components, population, labor force, and the like. First and most obvious are high growth rates of GDP per capita

and population in developed countries – both large multiples of the previous rates observable in these countries and of those in the rest of the world, at least until the recent decade or two» [1]. The trends that Simon Kuznets had identified for the two recent decades have become dominant in the contemporary world. The first one is a decline in population growth rates in devel-

oped countries down to depopulation. Such dynamics has been affecting the economic growth environment. To analyze the dynamics it was necessary to construct a global demographic balance.

In part 2 of this paper the global demographic balance methodology is presented. It's possible to estimate quantitative parameters of the global demographic situation with the help of the global demographic balance (the respective estimates are provided in the third part of this article). The quality characteristics of countries of the world have been assessed by applying the HDI, the findings are presented in the fourth part of this paper. Special attention has been given to the situation in BRICS countries, including Russia.

## 2. Global demographic balance methodology

To build the global demographic balance data from «World Populations Prospects. The 2008 Revision. Volume II: Sex and Age Distribution of the World Populations» [2] have been used.

Values for the following five-year age cohorts have been calculated:

- total population;
- children (0-14);
- persons of working age (15-59);
- retirement-age population (60 or over);
- the extreme elderly (80 or over).

For each population group its percentage of total population both of a region and the whole world has been calculated. In addition, the annual growth rate for each interval value indicator ( $r$ ) has been determined:

$$r = \frac{\ln P_t - \ln P_0}{t} \quad (1)$$

Where:

$P_t$  – the value of the index at the end of an interval;

$P_0$  – the value of the index at the beginning of an interval;

$t$  – the length of an interval.

To analyze the global demographic balance the following regionalization structure has been chosen to classify countries and regions of the world:

1. World as a whole;
2. More developed countries;
3. Less developed countries;
4. Europe;
5. Western Europe;

6. Eastern Europe;
7. Russia;
8. North America;
9. USA;
10. Latin America;
11. Brazil;
12. Oceania;
13. Asia;
14. China;
15. India;
16. Japan;
17. Africa;
18. Sub-Saharan Africa;
19. South Africa;
20. BRICS.

The global demographic balance has enabled to analyze core demographic trends in the world, in civilizations and in retrospective according to the UN average and highest projections.

## 3. Analysis of world demographic trends with the global demographic balance

On the planet rapid population growth is observed in a high period of industrial civilization. Previous population declines were results of wars, epidemics and natural disasters, but after such periods the population started to rise again. The crisis of industrialization and the approach of post-industrial civilization called into being a new demographic phenomenon.

1. First of all, one can observe a demographic change associated with the decline in birth rates and population ageing in developed countries (representing West European, East European, North American civilizations, a part of Eurasian one, as well as Australia and New Zealand from Oceania). Such demographic processes correlate closely with the Human Development Index and the level of a country development. The consumer society formed in the most developed countries (that comprised 1 billion 273 million or accounted for 18% of the world population in 2010) has stopped reproducing and requires more and more resources to support a prolonged lifetime. In 1950–2000 the increase in the number of children (0-14 years old) had stabilized in the above listed countries, whereas in 2050 the number of children is expected to fall to 10%, and only 11% of children will live from their birth in the consumer society, according to the UN average projections. At the

same time, the number of retirees (60+) had increased by 2.4 times in the mentioned countries in 1950–2000, and in 2000–2050 it is estimated to rise by 3.1 times. The number of the extreme elderly had grown by 4.4 times in 1950–2000, and in 2000–2050 it will augment by 3.2 times to reach 9% of the total population.

The number of persons of working age (15–59) had increased by 1.5 times in the most developed countries in 1950–2000, but it will decline by 11% in 2000–2050 to 52% of the total population, according to the UN average projections. It implies that only every eighth inhabitant of the planet will enjoy postindustrial production and consumption patterns.

In Western Europe, representing the western civilization, according to the average projections, the average population growth rate is estimated to drop from 0.7% in 1950–1975 to 0.15% in 2030–2050 that will constitute 2.7% of the total population. Such a demographic load on the population of working age implies extension of the working age among the population (60+) thanks to higher public healthcare level and lifelong education. Besides, postindustrial production tends to reduce the number of persons employed in the material production.

Perhaps, it will be followed by reductions of persons employed in human services along with virtualization of this area. Besides, possibly, immigrants from Asia and Africa will continue to move to Western Europe to be involved in unskilled work, and immigrants from Eastern Europe and Eurasia – in skilled work.

Eastern Europe (representing East European and Eurasian civilizations) has faced depopulation not because of demographic changes but due to social upheavals generated by the dissolution of the socialist camp and the USSR. A sudden fall in income and healthcare spending due to transition from socialistic principles to the market economy had resulted in the decline in population: (-0.4%) in 2000–2010, with a predictable fall to (-0.1%) – (-0.5%), accompanied by slowdown in children growth to (-0.8%) in 1975–2000 and to (-2.5%) in 2000–2010. That's why countries of Eastern Europe should pursue a purposeful policy to support birth rates and social stability.

In the USA (representing North American civilization) high population growth rates are observed with positive children growth rates over the entire interval, as well as shrinking share of the employed (15–59) – to 55% and higher percentage of the elderly (80+) – to 7–8%.

In Japan (representing Japanese civilization) the population growth rates are estimated at (-0.17%) –

(-0.39%) for 2010–2030 and (-0.36%) – (-0.72%) for 2030–2050. Percentage of children will drop to 11–15%, and the percentage of employed people – to 45%. At the same time, percentage of retirees (60+) will rise to 40–44%, and percentage of the extreme elderly – to 14–15%. Such demographic levels will force the Japanese society to identify new ways to handle this challenge.

Such demographic processes in developed countries will cause transformations in the economy and culture. The capitalist system had been operating, first of all, during a long historical period of demographic expansion. Nowadays it needs to adjust for a period, when population with high potential has been shrinking. It has to modify mechanisms to skim extra profit from world economic streams for the Golden Billion population that is aging. Besides, this process may impact the dominant system of sociocultural values. As different age groups have different values, educational and cultural interests, the sociocultural sphere of developed countries will more and more represent interests of the mature age population of those countries when the virtual component of this sphere will generally expand.

2. The least developed countries of the world (49 countries representing African, Muslim and Buddhist civilizations) accounted for 12% of the total population or 854 million in 2010. In the above mentioned countries the population had grown by 3.4 in 1950–2000 and is expected to increase by 2.5 in 2050, according to the UN average projections, and by 2.8 according to the highest estimates. These countries will face a slowdown in growth rates from 3.8 times in 1950–2000 to 1.5 times in 2000–2050, according to the UN average projections, and to 2 times according to the highest ones. At the same time, while the growth rate of retirees slightly lagged behind the population growth rate in the mentioned countries in 1950–2000, in 2000–2050 it will augment to 5.5 and 7.9 for the extreme elderly (80+). Consequently, in 2050 every fourth child, every 20th retiree and every fifth employed habitant of the planet will live in 49 least developed countries of the world.

The Human Development Index is extremely low in these countries. Correspondingly, every fourth child of the planet will live under conditions, where educational and healthcare services are almost unavailable for him/her. The same situation will be faced by every fifth employee in the world. As a result, the fifth part of human potential isn't good for functioning in the postindustrial civilization; the main field of using their work is old technological structures. According to the UN data, at present 1 billion people starve on the planet. Humaniza-

tion of life on the basis of partnership of civilizations assumes intensification of redistribution processes in favor of poor habitants in the world. It's feasible only due to robust growth in labor productivity, adoption of the sixth technological structure in developed and developing countries of the world with simultaneous intensification of redistribution streams from developed and developing countries to the poorest ones.

3. Demographic processes in developing countries maintain a middle position between the most and the least developed countries. Let's examine characteristics of these processes using BRICS countries (representing Latin American, Eurasian, Indian, Chinese and African civilizations), as an example.

The population of BRICS accounted for 44% of the world population in 1975 and is projected to shrink to 37% in 2050, while enjoying positive population growth that is estimated to decrease from 2% in 1950–1975 to 0.1% in 2050. In 2010 39% of children of the world, 45% of the employed, 40% of the retired and 33% of the extreme elderly (80+) lived in the region. In 2050 children will account for 17–21%, the employed – 57–58%, the retired – 22–25% and the extreme elderly – 4–5% of the total population of the mentioned countries.

Russia, as a part of BRICS, represents Eurasian civilization, and it is one of the most developed countries of BRICS. This country saw depopulation (-0.4%) in 2000–2010, and depopulation rates are expected to persist at (-0.2%) – (-0.4%) in 2010–2030 and at (-0.1%) – (-0.5%) in 2030–2050. In 1975–2000 the increase in the number of children averaged (-0.6%), in 2000–2010 – (-2.4%). At the same time, in 2000–2010 the country faced the decrease in the number of retirees (-0.6%). This information is an illustration of declined birth rates and higher mortality rates caused by the social catastrophe in the 1990s. As a result, Russia's percentage of the world population will shrink from 4% in 1950 to 1.2% in 2050. National demographic policies aimed to encourage births and to decrease mortality have started to produce positive results but the country is in the period of the demographic echo of the birth decline of the 1990s that aggravates the situation.

#### 4. Use of the Human Development Index to estimate the world demographic situation

In the previous part of this paper the quantitative demographic indices of countries of the world were examined. The HDI use enables to assess quality indices associated with development of countries of the world.

Since the 1990s «the Washington Consensus» has been governing the world. But some politicians in developing countries have formulated their own purposes and development mechanisms. First of all, these include Pakistani Mahbub-ul-Haq and Indian Amartya Sen. In 1990 they suggested a new integral indicator to assess national progress. That was the Human Development Index (HDI). They assumed that «the real wealth is people» not profits earned by those people.

Since 1990 the United Nations Development Programme has published series of reports focusing on cross-country dimension of the HDI.

Experiences of 20-year operations in this field were summarized in Human Development Report 2010 [3].

The HDI constitutes a new pattern of progress that consists of three components: Gross National Income per capita, life expectancy and educational level. At the same time, the report stipulated that «human development is the expansion of people's freedoms to live long, healthy and creative lives; to advance other goals they have reason to value; and to engage actively in shaping development equitably and sustainably on a shared planet. People are both beneficiaries and the drivers of human development, as individuals and in groups» [3, p. 2].

The world's average HDI has increased 18% since 1990 and 41% since 1970. Three of 135 countries – the Democratic Republic of the Congo, Zambia and Zimbabwe have a lower HDI today than in 1970. The slowest progress has been experienced by countries in Sub-Saharan Africa struck by the HIV epidemic and countries in the former Soviet Union suffering increased adult mortality. The top HDI movers include China, Indonesia and South Korea (income «growth miracles»), as well as Nepal, Oman and Tunis in other categories. The slowdown in health advances is due largely to dramatic reversals in 19 countries. In 9 of them – 6 in Sub-Saharan Africa and 3 in the former Soviet Union – life expectancy has fallen below 1970 levels [3, pp. 3-4].

The Human Development Report specifies two central contentions: that human development is different from economic growth and that substantial achievements are possible even without fast growth [3, p. 5].

The report emphasizes the lack of a significant correlation between economic growth and improvements in health and education. Research by William Easterly and François Bourguignon had confirmed that. «Iran, Togo and Venezuela experienced income declines, yet their life expectancy has risen an average of 14 years and their gross school enrolment an average of 31%

since 1970 [3, p. 47]. This conclusion is very valuable for Russia, especially in the period of international sanctions against it.

The report stipulates that «much development policy-making assumes that economic growth is indispensable to achievements in health and education. Our results suggest that this is not the case. This does not mean that countries can forget about growth – we have underlined that growth generates important possibilities. Rather, the results imply that countries do not have to solve the difficult problem of generating growth in order to tackle many problems of the health and education fronts» [3, p. 50].

The stable development generates opportunities to maintain this positive process in the future. The HDI estimates this opportunity quite comprehensively. Not only the dynamics of economic growth but also both the educational level and life expectancy reached by this generation constitute the basis of high life quality for next generations. Especially progress in health confirms that. During half a year life expectancy has increased in developing countries as much as in the developed ones. The epidemiological transition has been passed. Some 85% of mortality reductions in a sample of 68 countries since 1950 could be explained by global progress. Eradicating smallpox, a disease that killed some 2 million people annually in the 1960s, cost only \$300 million.

Concerted government policies mattered. A package of six vaccines assembled by World Health Organization costs less than \$1, and deworming costs just 50 cents a year [3, pp. 50-51]. That has reaffirmed the conclusion made above about close correlation between economic growth and achievements in health and education in developing countries in the period of global progress.

The *table 1* lists HDI values for BRICS countries, according to Human Development Report 2010.

As the figures in the *table 1* show, BRICS countries don't have high indices in terms of the HDI standard [3, pp. 143-145]. Norway (HDI 0.938), Australia (0.937), New Zealand (0.907), USA (0.902) and Ireland (0.895) are top 5 HDI countries [3, p. 143].

As far as BRICS countries are concerned, Russia demonstrates the best index (0.719) that corresponds to rank 65, and India has the worst one (0.519) with rank 119. Such low HRI values of BRICS countries suggest that the quality of human potential in these countries has sound growth prospects.

### 5. Conclusions

According to the UN average estimate (disregarding the biosphere ultimate capacity), the world population is forecast to reach 9 billion 150 million, and according to the highest estimate, it may rise to 10,461 million. According to S.P. Kapitsa's [4] projection, the world population may approach 11,300 million people due to hyperbolic growth and demographic change. This projection is based on absolutely strong growth, ignoring any limitations associated with the biosphere capacity. At the same time, the ultimate earthen biosphere capacity is 5 billion people, as G.P. Gorshkov's [5] calculations suggest. According to B.M. Dolgopолоv's projection [6], the total population will stabilize at around 5.6 billion people with allowance for the biosphere capacity. A.A. Akayev's [7] forecasts an aperiodic or oscillatory reversal of the world population to about 5.2 billion people. U.V. Yakovets [8] supposes that the technological progress removes the limitations of the ultimate biosphere capacity. K.E. Tsiolkovsky [9] shares this opinion.

Basically, the UN demographic projections disregard the ultimate biosphere capacity. According to these projections, the total population had grown by 2.4 times in 1950 – 2000 and is estimated to rise

Table 1.

The Human Development Index (HDI) in BRICS countries

Country	HDI rank	HDI value	Life expectancy at birth (years)	Mean years of schooling	Gross National Income (GNI) per capita (PPP 2008 \$)
Russia	65	0.719	67.2	8.8	15,258
Brazil	73	0.699	72.9	7.2	10,607
China	89	0.663	73.5	7.5	7,258
South Africa	110	0.597	52.0	8.2	9,812
India	119	0.519	64.4	4.4	3,337

by 1.5–1.7 times further in 2000–2050. The authors have relied on these projections to compute their own global demographic balance. The following results have been obtained: in 2000–2050 the total number of children is expected to grow by 1–1.3 times, the employed – by 1.5–1.6 times, the retired – by 3 times and the extreme elderly (80+) – by 5.6 times. According to these calculations, children will account for 20–24% of the world population, the employed – 57–58%, the retired – 19–22% and the extreme elderly (80+) – 4% by 2050.

Overall, population aging is observed because of the decline in births and the increase of the number of re-

tirees, especially the extreme elderly (80+). That confirms Kapitsa's conclusion that more and more countries of the world are facing demographic changes. The HDI analysis made by the authors to compare BRICS with the most developed countries has enabled them to draw a conclusion about sound growth potential for BRICS countries, including Russia.

Holding the demographic imperative affects production and consumption patterns. Correspondingly, in the middle of the 21<sup>st</sup> century serious changes are anticipated in the global socio-economic landscape and in economic growth environment that necessitates new research in this area. ■

### References

1. Kuznets S. (1971) Nobel Prize lecture. Available at: [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1971/kuznets-lecture.html](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1971/kuznets-lecture.html) (accessed 01 November 2014).
2. United Nations (2009) *World Populations Prospects. The 2008 Revision. Volume II: Sex and age distribution of the world populations*. NY: United Nations.
3. United Nations (2010) *Doklad o razvitiu cheloveka 2010. Real'noe bogatstvo narodov: puti k razvitiu cheloveka* [Report on development of the person 2010. Real wealth of the people: ways to development of the person]. Available at: <http://www.un.org/ru/development/hdr/2010/> (accessed 01 November 2014). (in Russian)
4. Kapitsa S.P. (1999) *Obshhaja teorija rosta chelovechestva* [General theory of growth of mankind]. M.: Nauka. (in Russian)
5. Gorshkov V.G. (1995) *Fizicheskie i jekologicheskie osnovy ustojchivosti zhizni* [Physical and ecological bases of stability of life]. Edited by K.S.Losev. M.: VINITI. (in Russian)
6. Korotaev A.V., Grinin L.E., Malkov S.Ju., Eds. (2010) *Analiz i modelirovanie global'noj dinamiki* [Analysis and modeling of global dynamics]. M.: Librukom. (in Russian)
7. Akaev A.A., Korotaev A.V., Malinetsky G.G., Malkov S.Ju., Eds. (2010) *Proekty i riski budushhego. Konceptii. Modeli. Instrumenty. Prognozy* [Projects and risks of the future. Concepts. Models. Tools. Forecasts]. Moscow: Jeditorial URSS. (in Russian)
8. Jakovets Ju.V., Abramov V.L. (2012) *Analiz faktorov nauchno-tehnologicheskogo razvitija v kontekste civilizacionnyh ciklov* [The analysis of factors of scientific and technological development in the context of civilization cycles]. Moscow: MISK. (in Russian)
9. Tsiolkovsky K.E. (1986) *Grezy o Zemle i nebe* [Dreams about Earth and the sky]. Tula: Priokskoe knizhnoe izdatel'stvo. (in Russian)

# МИРОВАЯ ДЕМОГРАФИЧЕСКАЯ СИТУАЦИЯ В РАКУРСЕ ГЛОБАЛЬНОГО ДЕМОГРАФИЧЕСКОГО БАЛАНСА

## **Т.К. БОГДАНОВА**

кандидат экономических наук, доцент кафедры бизнес-аналитики, факультет бизнес-информатики, Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: tanbog@hse.ru

## **Т.Ю. ЯКОВЕЦ**

кандидат экономических наук, член-корреспондент Российской академии естественных наук, докторант Института социально-экономических проблем народонаселения Российской академии наук

Адрес: 117218, г. Москва, Нахимовский пр-т, д. 32

E-mail: tzag@mail.ru

В 1971 в Нобелевской лекции С. Кузнец отметил, что за последние 1–2 десятилетия рост населения перестает быть основными двигателем экономического роста. В связи с этим авторами исследуется современная демографическая ситуация в мире на базе информации, приведенной в прогнозах ООН. Описывается методология глобального демографического баланса, включающего пять возрастных когорт населения для 20 стран и регионов мира для прошлого периода, для 1950–2010 гг. и прогноза ООН до 2050 года. С его помощью количественно анализируется демографическая ситуация в развитых, наименее развитых и развивающихся странах мира. Для развитых стран мира, прошедших демографический переход, в XXI веке будет характерен режим депопуляции населения. Приводится возрастная структура депопуляционных тенденций в этих странах. В наименее развитых странах рост населения продолжается, но уже не такими высокими темпами, как во второй половине XX века. Для развивающихся стран особо выделяются страны БРИКС. Для оценки качественных характеристик стран мира используется Индекс развития человеческого потенциала (ИРЧП). Даются характеристики данного индикатора, приводимые Программой развития ООН. Приводятся значения ИРЧП для стран БРИКС и делается вывод о необходимости качественного роста экономик этих стран. Приводятся результаты моделирования и прогнозирования Г.П. Горшковым, Б.М. Долгополовым и А.А. Акаевым численности населения мира с учетом предельной емкости биосферы. Делается вывод о том, что прогнозы С.П. Капицы и экспертов ООН, сделанные без учета предельной емкости биосферы, более реалистичны.

**Ключевые слова:** демографический прогноз ООН, глобальный демографический баланс, возрастная структура населения, развитые, наименее развитые и развивающиеся страны мира, страны БРИКС, индекс развития человеческого потенциала, Программа развития ООН, предельная емкость биосферы.

## **Литература**

1. Kuznets S. Nobel Prize lecture. [Электронный ресурс]: [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1971/kuznets-lecture.html](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1971/kuznets-lecture.html) (дата обращения 01.11.2014).
2. World Populations Prospects. The 2008 Revision. Volume II: Sex and age distribution of the world populations. New York: United Nations, 2009. 963 p.
3. Доклад о развитии человека 2010. Реальное богатство народов: пути к развитию человека. [Электронный ресурс]: <http://www.un.org/ru/development/hdr/2010/> (дата обращения 01.11.2014).
4. Капица С.П. Общая теория роста человечества. М.: Наука, 1999. 117 с.
5. Горшков В.Г. Физические и экологические основы устойчивости жизни /Отв. ред. К.С. Лосев. М.: ВИНТИ, 1995. 470 с.
6. Анализ и моделирование глобальной динамики / Под ред. А.В. Коротаева, Л.Е. Гринина, С.Ю. Малкова. М.: Либруком, 2010. 352 с.
7. Проекты и риски будущего. Концепции. Модели. Инструменты. Прогнозы /Под ред. А.А. Акаева, А.В. Коротаева, Г.Г. Малинецкого, С.Ю. Малкова. М.: Эдиториал УРСС, 2010. 432 с.
8. Анализ факторов научно-технологического развития в контексте цивилизационных циклов /Под ред. Ю.В. Яковца, В.Л. Абрамова. М.: МИСК, 2012. 456 с.
9. Циолковский К.Э. Грезы о Земле и небе. Тула: Приокское книжное издательство, 1986. 264 с.

Представляемая для публикации статья должна быть актуальной, обладать новизной, отражать постановку задачи (проблемы), описание основных результатов исследования, выводы, а также соответствовать указанным ниже правилам оформления.

Текст должен быть тщательно вычитан автором, который несет ответственность за научно-теоретический уровень публикуемого материала.

Материалы представляются в электронном виде по адресу:  
bijournal@hse.ru.

## ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

**ТЕКСТ СТАТЬИ** представляется в редакцию в электронном виде (в формате MS Word, версия 2003 или выше).

**ОБЪЕМ.** Ориентировочный объем статьи составляет 20-25 тысяч знаков (с пробелами).

### ШРИФТ, ФОРМАТИРОВАНИЕ, НУМЕРАЦИЯ СТРАНИЦ

**ШРИФТ** – Times New Roman, кегль набора – 12 пунктов, полуторный интервал, форматирование по ширине. Нумерация страниц – сверху по центру, поля: левое – 2,5 см, верхнее, нижнее и правое – по 1,5 см.

**НАЗВАНИЕ СТАТЬИ** приводится на русском и английском языках. Название статьи должно быть информативным и раскрывать содержание статьи.

**СВЕДЕНИЯ ОБ АВТОРАХ** приводятся на русском и английском языках и включают следующие элементы:

- ◆ фамилия, имя, отчество всех авторов полностью
- ◆ должность, звание, ученая степень каждого автора
- ◆ полное название организации – места работы каждого автора в именительном падеже, полный почтовый адрес каждой организации (включая почтовый индекс)
- ◆ адрес электронной почты каждого автора.

**АННОТАЦИЯ К СТАТЬЕ** представляется на русском и английском языках.

- ◆ Объем – 200-300 слов.
- ◆ Аннотация должна быть информативной (не содержать общих слов).
- ◆ Аннотация должна отражать основное содержание статьи и быть структурированной (следовать логике описания результатов в статье).
- ◆ Структура аннотации: предмет, цель, метод или методологию проведения исследования, результаты исследований, область их применения, выводы.
- ◆ Метод или методологию проведения исследований целесообразно описывать в том случае, если они отличаются новизной или представляют интерес с точки зрения данной работы. В аннотациях статей, описывающих экспериментальные работы, указывают источники данных и характер их обработки.
- ◆ Результаты работы описывают предельно точно и информативно. Приводятся основные теоретические и экспериментальные результаты, фактические данные, обнаруженные взаимосвязи и закономерности. При этом отдается предпочтение новым результатам и дан-

ным долгосрочного значения, важным открытиям, выводам, которые опровергают существующие теории, а также информации, которая, по мнению автора, имеет практическое значение.

- ◆ Выводы могут сопровождаться рекомендациями, оценками, предложениями, гипотезами, описанными в статье.
- ◆ Сведения, содержащиеся в названии статьи, не должны повторяться в тексте аннотации. Следует избегать лишних вводных фраз (например, «автор статьи рассматривает...»).
- ◆ Исторические справки, если они не составляют основное содержание документа, описание ранее опубликованных работ и общеизвестные положения, в аннотации не приводятся.
- ◆ В тексте аннотации следует употреблять синтаксические конструкции, свойственные языку научных и технических документов, избегать сложных грамматических конструкций.
- ◆ В тексте аннотации следует применять значимые слова из текста статьи.

**КЛЮЧЕВЫЕ СЛОВА** приводятся на русском и английском языках. Количество ключевых слов (словосочетаний) – 6-10. Ключевые слова или словосочетания отделяются друг от друга точкой с запятой.

**ФОРМУЛЫ.** При наборе формул, как выключных, так и строчных, должен быть использован редактор формул Math Type или MS Equation. В формульных и символических записях греческие (русские) символы, а также математические функции записываются прямыми шрифтами, а переменные аргументы функций в виде английских (латинских) букв – наклонным курсивом (пример «cos a», «sin b», «min», «max»). Нумерация формул – сквозная (по желанию авторов допускается двойная нумерация формул с указанием структурного номера раздела статьи и, через точку, номера формулы в разделе).

**РИСУНКИ** (графики, диаграммы и т.п.) могут быть оформлены средствами MS Word или MS Excel. Ссылки на рисунки в тексте обязательны и должны предшествовать позиции размещения рисунка. Допускается использование графического векторного файла в формате wmf/emf или cdr v.10. Фотографические материалы предоставляются в формате TIF или JPEG, с разрешением изображения не менее 300 точек на дюйм. Нумерация рисунков – сквозная.

**ТАБЛИЦЫ** оформляются средствами MS Word или MS Excel. Нумерация таблиц – сквозная.

**СПИСОК ЛИТЕРАТУРЫ** составляется в соответствии с требованиями ГОСТ 7.0.5-2008. Библиографическая ссылка (примеры оформления размещены на сайте журнала <http://bi.hse.ru/>). Нумерация библиографических источников – в порядке цитирования. Ссылки на иностранную литературу – на языке оригинала без сокращений.

**СПИСОК ЛИТЕРАТУРЫ ДЛЯ АНГЛОЯЗЫЧНОГО БЛОКА** оформляется в соответствии с требованиями SCOPUS (примеры оформления размещены на сайте журнала <http://bi.hse.ru/>). Для транслитерации русскоязычных наименований можно воспользоваться сервисом <http://translit.ru/>.

## ЛИЦЕНЗИОННЫЙ ДОГОВОР

Для размещения полнотекстовых версий статей на сайте журнала с авторами заключается лицензионный договор о передаче авторских прав.

Плата с авторов за публикацию рукописей не взимается.

# AUTHORS GUIDELINES

Articles should be topical and original, should outline tasks (issues), describe key results of the author's research and appropriate conclusions.

Manuscripts are submitted via e-mail: [bijournal@hse.ru](mailto:bijournal@hse.ru).

## MANUSCRIPT REQUIREMENTS

**TEXT FILES** should be submitted in electronic form, as a MS Word document (version 2003 or higher).

**LENGTH.** Articles should be between 20 and 25 thousand characters (incl. spaces).

**FONT, SPACING, MARGINS.** The text should be in Times New Roman 12 pt, 1.5 spaced, fit to the width, margins: left – 25 mm, all other – 15 mm.

**TITLE** of the article should be submitted in native language and English.

**AUTHORS' DETAILS** are presented in native language and English. The details include:

- ◆ Full name of each author
- ◆ Position, rank, academic degree of each author
- ◆ Affiliation of each author, at the time the research was completed
- ◆ Full postal address of each affiliation (incl. postcode / ZIP)
- ◆ E-mail address of each author.

**ABSTRACT** are presented in native language and English.

- ◆ The abstract should be between 200 and 300 words.
- ◆ The abstract should be informative (no general words), original, relevant (reflects your paper's key content and research findings); structured (follows the logics of results' presentation in the paper)
  - ◆ The recommended structure: purpose (mandatory), design / methodology / approach (mandatory), findings (mandatory), research limitations / implications (if applicable), practical implications (if applicable), originality / value (mandatory).
- ◆ It is appropriate to describe the research methods/methodology if they are original or of interest for this particular research. For

papers concerned with experimental work the data sources and data procession technique should be described.

- ◆ The results should be described as precisely and informatively as possible. Include your key theoretical and experimental results, factual information, revealed interconnections and patterns. Give special priority in the abstract to new results and long-term impact data, important discoveries and verified findings that contradict previous theories as well as data that you think have practical value.
- ◆ Conclusions may be associated with recommendations, estimates, suggestions, hypotheses described in the paper.
- ◆ Information contained in the title should not be duplicated in the abstract. Authors should try to avoid unnecessary introductory phrases (e.g. «the author of the paper considers...»).
- ◆ Authors should use the language typical of research and technical documents to compile your abstract and avoid complex grammatical constructions.
- ◆ The text of the abstract should include key words of the paper.

**KEYWORDS** are presented in native language and English. The number of key words / words combinations are from 6 to 10 (separated by semicolons).

**FORMULAE** should be prepared using Math Type or MS Equation tool.

**FIG.S** should be of high quality, black and white, legible and numbered consecutively with Arabic numerals. All Fig.s (charts, diagrams, etc.) should be submitted in electronic form (photo images – in TIF, PSD or JPEG formats, minimum resolution 300 dpi). Appropriate references in the text are required.

**REFERENCES** should be presented in Harvard style and carefully checked for completeness, accuracy and consistency.

The publication is free of charge.