Бизнес-ИНФОРМАТИКА

МЕЖДИСЦИПЛИНАРНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

СОДЕРЖАНИЕ

О ЖУРНАЛЕ

«Бизнес-информатика» — рецензируемый междисциплинарный научный журнал, выпускаемый с 2007 года Национальным исследовательским университетом «Высшая школа экономики» (НИУ ВШЭ). Администрирование журнала осуществляется школой 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:

✦ анализ данных и интеллектуальные системы

✦ информационные системы и технологии в бизнесе

✦ математические методы и алгоритмы бизнес-информатики

✦ программная инженерия

✦ Интернет-технологии

✦ моделирование и анализ бизнес-процессов

✦ стандартизация, сертификация, качество, инновации

✦ правовые вопросы бизнес-информатики

✦ принятие решений и бизнес-интеллект

✦ моделирование социальных и экономических систем

✦ информационная безопасность.

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.

BUSINESS INFORMATICS

INTERDISCIPLINARY ACADEMIC JOURNAL

CONTENTS

ABOUT
THE JOURNAL

Business 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 School 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:

✦ data analysis and intelligence systems

✦ information systems and technologies in business

✦ mathematical methods and algorithms of business informatics

✦ software engineering

✦ Internet technologies

✦ business processes modeling and analysis

✦ standardization, certification, quality, innovations

✦ legal aspects of business informatics

✦ decision making and business intelligence

✦ modeling of social and economic systems

✦ information security.

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.

INTERNET GOVERNANCE: TRENDS AND REALITIES. PART 1

Peter MAJOR

*Vice-Chairman, United Nations Commission on Science
and Technology for Development (CSTD)*

Address: Palais des Nations, 8-14, Av. de la Paix, 1211 Geneva 10, Switzerland
E-mail: pmajor@bluewin.ch

Internet Governance is one of the global governance issues that emerged at the end of the 1990s. As the Internet is taking on an important role in every aspect of our lives, it has been recognized that its governance needs to involve all stakeholders and institutions to allow global access, foster development and contribute to the global economy, education, information. Security and privacy concerns should also be addressed in the policy dialogues. This paper outlines the global nature of the Internet Governance. It argues that in addition to dialogues and negotiations formal approaches should be explored to handle global issues in a global way. In its conclusion a mechanism is proposed to address Internet related global public policy issues.

The views expressed in this paper are solely those of the author in his private capacity and do not in any way represent the views of the UN Commission on Science and Technology for Development.

Introduction

More than 3 billion people, 42% of the world's population had access to the Internet as of the second quarter of 2014 [1]. This is a 741% growth from 2000 to 2014. The Internet has evolved from an academic/military project to a global public utility [2].

The phenomenal evolution of the Internet raises many questions about its governance. Since the World Summit on the Information Society (WSIS) organized by the UN in December 2003 in Geneva and December 2005 in Tunis (*http://www.itu.int/wsis/*), many initiatives and proposals from governments, the technical community, academia, business, civil society and international organizations have tried to grasp the complexity of Internet Governance. The Geneva Plan of Action and Tunis Agenda documents reflected our understanding and assessment of the global public policy issues related to the Internet. At the end of 2015, the UN General Assembly will review the implementation of the Action Lines formulated in the Geneva Plan of Action and will decide on the way forward. It is useful to assess the achievements from a different angle.

In this paper, I analyze the global nature of Internet Governance, its common characteristics with other global governance systems, as well as its specific features. After dealing with governance, global issues and global governance approaches in general, I describe the characteristics of Internet Governance, the similarities we can find with other governance systems and the impact of the growth and global penetration of the Internet on IG.

There are many open questions: how can we extend off-line arrangements to the on-line space? are traditional national governance mechanisms adaptable to the Internet? and can we apply formal approaches to Internet Governance?

At the end of the paper, I briefly discuss the lessons learned and what is the way forward Internet Governance.

1. Governance, global issues, global governance

In traditional governance, a governing or political authority and institutions ultimately exercise control; formal political institutions aim at coordinating and controlling interdependent social relations that also possess the capacity to enforce decisions. Informal governance works more toward practices and guidelines. In the process of governance, collective interests are articulated, rights and obligations are established and differences are mediated.

A possible definition of governance is offered by Thomas G. Weiss and Ramesh Thakur: «All processes of governing, undertaken by a government, market or network, over a family, tribe, organization or territory through formal laws, norms, power or language or informal guidelines or practices» [3].

Issues such as climate change, nuclear arms proliferation, use of outer space or radio communication, to name a few, have trans-border or global impact. Handling these issues requires global approaches and ultimately global governance.

Global governance may be characterized as the «sum of laws, norms, policies, and institutions that define, constitute, and mediate trans-border relations between states, cultures, citizens, intergovernmental and non-governmental organizations, and the market. Global governance is the regulation of interdependent relations in the absence of overarching political authority, such as in the international system» [4].

Approaches to handling global issues reflect an appreciation of the problems and are influenced by the political environment. Up to the 1990s, global issues were mostly negotiated in a multilateral setting i.e.: in the country delegations only representatives of governments and some technical advisors participated in the discussions. Based on the results of negotiations, decisions were taken by the governments.

An early and successful example of global governance is the global regulatory framework of radio communication known as Radio Regulations (RR) [5]. The RR is an international treaty of the International Telecommunication Union (ITU), a specialized agency of the United Nations. The RR was created to avoid harmful interference. The first edition of the treaty dates from 1906. The text of the RR, considering rapid technological progress, needs to be modified regularly by the World Radiocommunication Conferences (WRC) organized by the ITU. In the preparatory phase and in the work of the Conference, stakeholders from the technical community,

business and academia participate as advisors within the respective national delegations. The signatories of the treaty are representatives of Member States.

More recent examples of multilateral global governance are the Outer Space Treaty (Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, 1967 [6]) and the International Treaty on the Non-proliferation of Nuclear Arms, 1968 [7].

Global governance is an ongoing process with successes and failures. Some of the failures may be attributed to governance gaps. There are five main governance gaps: knowledge gaps, normative gaps, policy gaps, institutional gaps and compliance gaps [4].

Knowledge gaps

There is often little or no consensus about the nature, causes, gravity, and magnitude of a problem, as well as about the empirical information or the theoretical explanation. There is often disagreement over the best remedies and solutions to these problems (examples: climate change, nuclear energy). The academic and technical community, such as universities, research institutes, scientific experts and think tanks, as well as civil society actors (NGOs) currently are playing a growing role in filling the gap.

Normative gaps

The norm is statistically a pattern of behavior that is most common or usual (normal curve), whereas an ethical norm is a pattern of behavior that should be followed in accordance with a given value system or the moral code of a society, a generally accepted standard of proper behavior. The two meanings may converge in practice; in most cases, they will complement each other; but in some cases, they may diverge. The UN is an arena where member states codify norms either as soft law (UN resolutions, declarations) or hard law (international conventions and treaties) seeking consensus about normative approaches to address global challenges. In practice, however, because of dissent by powerful states or mischief by large coalitions of less powerful ones, either no action occurs, or agreement is possible only on a lowest-common denominator.

Policy gaps

Policy is an interlinked set of governing principles and goals, and the agreed programs of action to implement those principles and achieve those goals. It is important

to understand that even if the source and scale of most of today's pressing challenges are global, and any effective solution to them must also be global, the policy authority for tackling these challenges remains vested in states. UN policy implementation is not by the Secretariat but by member states.

Institutional gaps

Institutions are formal organizations, but they may also be informal entities. If the policy is to escape the trap of being ad hoc, episodic, judgmental and idiosyncratic, it must be housed within an institution with resources and autonomy. Institutional gaps may exist for different reasons: absence of institutions, missing stakeholders or inadequate resources. It is up to the governments who are the driving force to fill the gap.

Compliance gaps

Agreed elements of international policy may not be complied with for lack of ability, resources or will of implementation. There is no clear mandate of any UN institution to have the monitoring authority, responsibility and capacity. Non-compliance may be defined as the lack of strength of conviction or commonality of interests to enforce the community norm. In the UN system, there is only a limited possibility and capacity to ensure that states comply with agreed international policy.

2. Global governance: multilateral approach

Interrelations of global issues have been explored in a systematic and scientific way ever since the second half of the last century. One of the most influential publications was «The Limits to Growth», commissioned by the Club of Rome in 1972 [8]. «The Limits to Growth» appeared at a time when human belief in the power of technology was at an all time high. There seemed to be no challenge that could not be overcome through application of human ingenuity and effort, in the form of economic growth based on continuing technological advance.

In this perspective, «The Limits to Growth» warned that the 2010-30 period would see some resources becoming scarce or expensive (for example regional scarcity of oil, water, fish, wood, land) while environmental damage would become increasingly visible (for example regional destruction of biodiversity, accumulation of toxics, ozone-destroying chemicals, and greenhouse gases). And importantly, all of this would occur in spite of continuing technological advance. «The limits to Growth»

emphasized that resource and pollution problems would occur because the world is physically finite — and actually quite small compared to the human footprint in the 21st century. The problems would start regionally, and gradually embrace the world — unless corrective action was taken immediately. Man was no longer omnipotent (*https://ic.fsc.org/*).

In response to global problems judged to be urgent, two more UN lead initiatives may be mentioned: the Montreal Protocol to protect the ozone layer (1989) [9] and the Kyoto Protocol, the UN Framework Convention on Climate Change (1992) [10].

The importance of global issues was fully recognized at the beginning of 1990s. The UN held its first Conference on Environment and Sustainable Development in Rio de Janeiro in 1992 and adopted an agenda for the environment and development in the 21st century. Agenda 21: A Programme of Action for Sustainable Development contains the Rio Declaration on Environment and Development, which recognizes each nation's right to pursue social and economic progress and assigned to States the responsibility for adopting a model of sustainable development. Agreements were also reached on the Convention on Biological Diversity and the Framework Convention on Climate Change [11]. This Rio conference was followed by the UN Conference on Population in 1994 in Cairo and subsequently by the UN Conference on Women in 1995, Beijing.

In 2000, the leaders of the world made a historic commitment: to eradicate extreme poverty and improve the health and welfare of the world's poorest people within 15 years. The commitment, adopted at the Millennium Summit in September 2000, was set forth in the United Nations Millennium Declaration [12]. This vision was expressed in eight time-bound goals, known as the Millennium Development Goals (MDGs):

✦ to eradicate extreme poverty and hunger;
✦ to achieve universal primary education;
✦ to promote gender equality and empower women;
✦ to reduce child mortality;
✦ to improve maternal health;
✦ to combat HIV/AIDS, malaria, and other diseases;
✦ to ensure environmental sustainability;
✦ to develop a global partnership for development.

High level review of the implementation of the Millennium Development Goals (2000 — 2015) and discussion on post-2015 goals (known as Sustainable Development Goals) held in September 2015, in the UN, New York.

3. Global governance: multi-stakeholder approach

In parallel with the full recognition of the importance of global issues, it has also been recognized that global issues should be addressed with the involvement of all stakeholders i.e.: with the representatives of governments, business, civil society, the technical and academic community and the international/intergovernmental organizations participating in the discussions and the decision-making. This approach is known as the multi-stakeholder approach.

The functions of multi-stakeholder initiatives are to identify global public needs, facilitate negotiation, gather and disseminate knowledge, create and correct markets, broaden participation in global governance.

The framework of multi-stakeholder initiatives needs international organizations, stakeholders need to be included early. The approach should be embedded in national policy debates. Transparency is instrumental and the approach should lead to accountability.

There are many challenges to multi-stakeholder initiatives: in addition to the known constraints of the multilateral approach influenced by divergent political interests of governments. Stakeholders from business have particular interests. Civil society has no clear legitimate representation and it is difficult to create mutual trust among different stakeholders considering that in the multi-stakeholder global governance model representatives of stakeholders participate in the discussions on an equal footing in their respective roles.

There are several examples of multi-stakeholder initiatives:

✧ The World Commission on Dams (WCD) formed in April 1997 to research the environmental, social and economic impacts of the development of large dams globally (*http://www.unep.org/DAMS/WCD/*). Members: civil society, academia, the private sector, professional associations and one government representative. Act in an individual capacity, not representing the organizations or governments of which they are members.

✧ The Forest Stewardship Council (FSC) is an international non-for-profit, multi-stakeholder organization established in 1993 to promote responsible management of the world's forests. Its main tools for achieving this are a standard setting, certification and labeling of forest products. The organization aims to «provide businesses and consumers with a ... tool to influence how forests worldwide are managed» (*https://ic.fsc.org/*).

✧ Global Fund to Fight Aids, Tuberculosis and Malaria Public-private partnership Programs are supported by the Global Fund in more than 140 countries. As of mid-2014, they had 6.6 million people on antiretroviral therapy for AIDS. They have tested and treated 11.9 million people for TB, and have distributed 410 million insecticide-treated nets to protect families against malaria (*http://www.theglobalfund.org/en/*).

4. Global Internet Governance

Similarly to other global issues, the relevance of ICTs and the global nature of the Internet were also recognized at the end of 1990s, leading to the WSIS. The notion of Internet Governance emerged after the first phase of the WSIS. The Declaration of Principles document adopted by the Geneva Summit [13] asked the Secretary General of the UN to establish a Working Group on Internet Governance «to investigate and make proposals for action, as appropriate, on the governance of the Internet by 2005». The task of this Working Group, established in 2004, was to organize an open dialogue on Internet Governance (WGIG) among all stakeholders and to bring recommendations on this subject to the second phase of the Summit.

The WGIG in its report proposed the following working definition of Internet governance: «Internet governance is the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the Internet» [14].

The WGIG identified the following public policy clusters:

(a) Issues relating to infrastructure and the management of critical Internet resources, including administration of the domain name system and Internet protocol addresses (IP addresses), administration of the root server system, technical standards, peering and interconnection, telecommunications infrastructure, including innovative and convergent technologies, as well as multilingualization. These issues are matters of direct relevance to Internet governance and fall within the ambit of existing organizations with responsibility for these matters;

(b) Issues relating to the use of the Internet, including spam, network security and cybercrime. While these issues are directly related to Internet governance, the nature of global cooperation required is not well defined;

(c) Issues relevant to the Internet have an impact much wider than the Internet and for which existing organizations are responsible, such as intellectual property rights (IPRs) or international trade. The WGIG started examining the extent to which these matters are being handled consistently with the Declaration of Principles;

(d) Issues relating to the developmental aspects of Internet governance, in particular capacity-building in developing countries.

During the second phase of WSIS held in Tunis in December 2005, the WGIG report served as the basis to create the Internet Governance Forum (IGF). As for the role of governments in Internet Governance to handle public policy issues, the magic formula of «enhanced cooperation» was introduced and discussions on implementation of enhanced cooperation are ongoing.

It is important to mention here some of the most significant organizations in the Internet Governance space.

The Internet Corporation for Assigned Names and Numbers (ICANN) (*https://www.icann.org/*) is a non-profit association (based in US since 1998). Its main mission is the management of identifiers (common pool resource). It is considered as a Global Multi-Stakeholder Agency with the following functions:

a. Notary (IANA functions and assignment of protocol parameters);

b. Competition Authority;

c. Regulator of the «Semantic Spectrum» (TLD space).

ICANN also develops and enforces policies (via contracts). Its governing body is an elected international Board (direct and ICANN Nominating Committee).

The Internet Society (ISOC), including the Internet Engineering Task Force (IETF) and the Internet Architecture Board (IAB), facilitates open development of standards, protocols, administration, and the technical infrastructure of the Internet. It supports education in developing countries specifically, and wherever the need exists it promotes professional development and builds a community to foster participation and leadership in areas important to the evolution of the Internet. It provides reliable information about the Internet.

The UN, ITU, UNESCO, UNCTAD, UNDP, UNEP, FAO, WIPO, WHO, WMO and other UN specialized agencies are facilitators for the WSIS Action Lines. The Commission on Science and Technology for Development (UN CSTD) is responsible for reviewing the activities of the facilitators for the WSIS Action Lines. There are several initiatives and forums, in addition to the formal organizations where different aspects of internet governance are discussed — NetMundial (*http://www.netmundial.br/*), Carl Bildt Initiative (*https://www.cigionline.org/activity/global-commission-internet-governance*), etc.

One of the proposals formulated in the WGIG report was to remove US government oversight of ICANN. This process started in March 2014 with the announcement of the US Government asking for proposals about the transition of its stewardship to a multi-stakeholder arrangement with a target date of 30 September 2015 (*https://www.icann.org/stewardship*).

Another proposal of the WGIG report was to create under the UN a discussion forum on Internet-related public policy issues with non-binding results. In 2006, the UN, following the outcomes of the WSIS, created a secretariat for the organization of the Internet Governance Forum (IGF). The first IGF was held in 2006 in Athens. The event concentrated on openness, critical Internet resources, security, diversity, emerging issues and the way forward. The IGF has become an annual event with an increasing number of participants during its 10 years of history. The IGF is a multi-stakeholder dialogue involving governments, civil society, academia and technical community, business and international organizations. It is not a decision making body, there are no formal negotiations and no binding documents are produced. It brings together stakeholders in an informal setting and creates a framework for cooperation. It also contributes to development and capacity building. Attendance of the IGFs varies from around 1000 (2006, Athens) to more than 3000 (2014, Istanbul). It is the task of the Multi-stakeholder Advisory Group nominated by the Secretary General of the UN to shape the IGF program, determining the overarching theme and the sub-themes, and to evaluate the workshop submissions. On average, there are more than 100 workshops under the sub-themes and several presentations organized by the dynamic coalitions (relatively informal, issue-specific groups consisting of stakeholders that are interested in the particular issue). In addition to workshops and dynamic coalition meetings, there are several activities taking place in the IGF meeting: best practices forums, open forums, flash sessions, inter-regional dialogue sessions, pre-events, and the IGF Village.

The main challenge of Internet Governance is to enable coexistence in shared cross-border digital spaces

The overarching themes and the main sessions of the IGFs

Year	Venue	Overarching theme	Sub-themes
2006	Athens (Greece)	Internet Governance for Development	• Security • Diversity • Access • The Way Forward • Emerging Issues
2007	Rio de Janeiro (Brazil)	Internet Governance for Development	• Critical Internet Resources • Access • Diversity • Security • Taking Stock and the Way Forward • Emerging Issues
2008	Hyderabad (India)	Internet for All	• Reaching the Next Billion, Promoting Cyber-Security and Trust • Managing Critical Internet Resources • Emerging Issues – the Internet of Tomorrow • Taking Stock and the Way Forward
2009	Sharm el Sheikh (Egypt)	Internet Governance – Creating Opportunities for All	• Critical Internet Resources • Security, Openness and Privacy • Diversity • Access • IG in the Light of WSIS Principles • Taking Stock • Emerging Issues – Impact of Social Networks
2010	Vilnius (Lithuania)	IGF 2010 – Developing The Future Together	• Managing Critical Internet Resources • Security, Openness and Privacy Access and Diversity • Internet Governance for Development • Emerging Issues: Cloud Computing Taking stock of Internet Governance and the Way Forward
2011	Nairobi (Kenya)	Internet as a Catalyst for Change: Access, Development, Freedoms and Innovation	• Internet Governance for Development • Emerging Issues • Managing Critical Internet Resources • Security, Openness and Privacy • Access and Diversity • Taking Stock and the Way Forward
2012	Baku (Azerbaijan)	Internet Governance for Sustainable Human, Economic and Social Development	• Emerging Issues • Managing Critical Internet Resources • Internet Governance for Development • Access and Diversity • Security, Openness and Privacy • Taking Stock and the Way Forward
2013	Bali (Indonesia)	Building Bridges – Enhancing Multi-stakeholder Cooperation for Growth and Sustainable Development	• Building Bridges – The Role of Governments in Multistakeholder Cooperation • Internet Governance Principles • Principles of Multistakeholder Cooperation • Legal and other Frameworks: Spam, Hacking and Cybercrime • (Access/Diversity): The Internet as an Engine For Growth and Sustainable Development • Human Rights, Freedom of Expression and Free Flow of Information on the Internet • Emerging Issues – Internet Surveillance • Open Microphone
2014	Istanbul (Turkey)	Connecting Continents for Enhanced Multistakeholder Internet Governance	• Policies Enabling Access • Content Creation, Dissemination and Use Internet as an Engine For Growth & Development • IGF & The Future of the Internet Ecosystem • Enhancing Digital Trust • Internet and Human Rights • Critical Internet Resources • Emerging Issues
2015	João Pessoa (Brazil)	Evolution of Internet Governance: Empowering Sustainable Development	• Cybersecurity and Trust • Internet Economy • Inclusiveness and Diversity • Openness • Enhancing Multistakeholder Cooperation • Internet and Human Rights • Critical Internet Resources • Emerging Issues

of billions of people with very diverse personal values and legal frameworks (*http://www.euro-ssig.eu/*). Recently the distinction has been proposed between governance of the Internet and governance on the Internet. In this approach, governance of the Internet covers the public policy cluster (a) identified by the WGIG and is formalized in the Internet protocol suite. Standards and protocols of the suite have been developed in a multi-stakeholder approach based on so-called rough consensus. Governance on the Internet covers public policy clusters (b), (c) and (d). The global nature of the Internet requires harmonization of legal systems, understanding of different political interests, acceptance of cultural differences, applying already existing legal frameworks to the Internet, etc. This leads to the need for identifying existing mechanisms dealing with global public policy issues related to the governance on the Internet and identifying gaps in the absence of these mechanisms.

The UN General Assembly in its resolution created the CSTD Working Group on Enhanced Cooperation (WGEC) with the mandate to «examine the mandate of WSIS regarding Enhanced Cooperation through seeking, compiling and reviewing inputs from all Member States and all other stakeholders, and to make recommendations on how to fully implement this mandate» [15]. The CSTD WGEC identified public policy issues and started a mapping exercise to identify existing mechanisms and gaps. The results of the exercise have been summarized in the conference room paper prepared by the CSTD Secretariat [16].

Discussions on the results of implementation of the Geneva Plan of Action and the Tunis Agenda have been conducted in UNESCO in Paris in 2013 followed by the High Level Meeting in the ITU in Geneva in 2014. The final high level review meeting of the WSIS will be held in December 2015 during the General Assembly of the United Nations in New York. ∎

References

1. Internet World Stats (2014) *Internet users in the world*. Available at: http://www.internetworldstats.com/stats.htm (accessed 31 January 2015).
2. Leiner B.M., Cerf V., Clark D.D., Kahn R.E., Kleinrock L., Lynch D.C., Postel J., Roberts L.G., Wolff S. (2015) A brief history of the Internet. *Internet Society*. Available at: http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet (accessed 24 February 2015).
3. Weiss T.G. (2013) *Global governance: Why? What? Whither?* Cambridge (UK): Polity Press.
4. Weiss T.G. (2009) The UN's role in Global Governance. *UN Intellectual History Project*, Briefing Note Number 15. Available at: http://www.unhistory.org/briefing/15GlobalGov.pdf (accessed 31 January 2015).
5. International Telecommunication Union (2012) *Radio regulations*. Geneva: ITU. Available at: http://www.itu.int/dms_pub/itu-s/oth/02/02/S02020000244501PDFE.pdf (accessed 31 January 2015).
6. United Nations Office for Outer Space Affairs (1967) *Treaty on principles governing the activities of states in the exploration and use of outer space*. Available at: http://www.unoosa.org/oosa/SpaceLaw/outerspt.html (accessed 31 January 2015).
7. United Nations Office for Disarmament Affairs (1968) *Treaty on the non-proliferation of nuclear weapons (NPT)*. Available at: http://www.un.org/disarmament/WMD/Nuclear/NPT.shtml (accessed 31 January 2015).
8. Club of Rome Organization (2012) *What was the message of «The limits to growth»*. Available at: http://www.clubofrome.org/flash/limits_to_growth.html (accessed 31 January 2015).
9. United Nations Environment Programme, Ozone Secretariat (1989) *The Montreal protocol on substance that depletes the ozone layer*. Available at: http://ozone.unep.org/new_site/en/montreal_protocol.php (accessed 31 January 2015).
10. United Nations Framework Convention on Climate Change (1992) *Kyoto Protocol*. Available at: http://unfccc.int/2860.php (accessed 31 January 2015).
11. United Nations Conference on Sustainable Development (2012) *The history of sustainable development in the United Nations*. Available at: http://www.uncsd2012.org/history.html (accessed 31 January 2015).
12. 12. United Nations (2000) *United Nations Millennium Declaration*. Available at: http://www.un.org/millennium/declaration/ares552e.htm (accessed 31 January 2015).
13. World Summit on the Information Society (2003) Document WSIS-03/GENEVA/DOC/4-E, 12 December 2003. Available at: http://www.itu.int/net/wsis/docs/geneva/official/dop.html (accessed 31 January 2015).
14. Working Group on Internet Governance (2005) *Report of the Working Group on Internet Governance, Chateau de Bossey, June 2005*. Available at: http://www.wgig.org/docs/WGIGREPORT.pdf (accessed 31 January 2015).
15. United Nations (2012) *United Nations General Assembly Resolution A/Res/67/195*. Available at: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/67/195 (accessed 31 January 2015).
16. United Nations CSTD Secretariat (2015) *The mapping of international Internet public policy issues. Conference room paper*. Geneva, Switzerland.

УПРАВЛЕНИЕ ИНТЕРНЕТОМ: ТЕНДЕНЦИИ И РЕАЛЬНОСТЬ. ЧАСТЬ 1

П. МЕЙОР
заместитель руководителя Комиссии ООН по науке и технологиям
Адрес: Palais des Nations, 8-14, Av. de la Paix, 1211 Geneva 10, Switzerland
E-mail: pmajor@bluewin.ch

Управление Интернетом представляет собой одно из направлений глобального управления, активно развивающееся с конца 1990-х годов. Поскольку Интернет играет важную роль в нашей жизни, представляется необходимым вовлечение в процессы управления всех заинтересованных лиц и институтов, а также способствование развитию и вкладу в мировую экономику, образование и информационное обеспечение. Вопросы информационной безопасности и конфиденциальности также должны рассматриваться в рамках обсуждения политики в области управления Интернетом. В данной статье рассматривается глобальный аспект управления Интернетом. Показано, что в дополнение к обсуждениям и переговорам должны применяться формальные подходы, позволяющие решать глобальные вопросы. Предлагается подход к решению вопросов, связанных с глобальной политикой в области управления Интернетом.

Выводы и рекомендации, приведенные в статье, отражают исключительно личное мнение автора и не должны трактоваться в качестве официальной позиции Комиссии ООН по науке и технологиям.

Литература

1. Internet users in the world (2014) // Internet World Stats [Электронный ресурс]: http://www.internetworldstats.com/stats.htm (дата обращения 31.01.2015).
2. A brief history of the Internet / B.M. Leiner [et al.] // Internet Society [Электронный ресурс]: http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet (дата обращения 24.02.2015).
3. Weiss T.G. Global governance: Why? What? Whither? Cambridge (UK): Polity Press, 2013. 180 pp.
4. Weiss T.G. The UN's role in Global Governance // UN Intellectual History Project, Briefing Note Number 15, 2009 [Электронный ресурс]: http://www.unhistory.org/briefing/15GlobalGov.pdf (дата обращения 31.01.2015).
5. Radio regulations. Geneva: International Telecommunication Union (ITU), 2012. [Электронный ресурс]: http://www.itu.int/dms_pub/itu-s/oth/02/02/S02020000244501PDFE.pdf (дата обращения 31.01.2015).
6. Treaty on principles governing the activities of states in the exploration and use of outer space. United Nations Office for Outer Space Affairs, 1967 [Электронный ресурс]: http://www.unoosa.org/oosa/SpaceLaw/outerspt.html (дата обращения 31.01.2015).
7. Treaty on the non-proliferation of nuclear weapons (NPT). United Nations Office for Disarmament Affairs, 1968 [Электронный ресурс]: http://www.un.org/disarmament/WMD/Nuclear/NPT.shtml (дата обращения 31.01.2015).
8. What was the message of «The limits to growth». Club of Rome Organization, 2012 [Электронный ресурс]: http://www.clubofrome.org/flash/limits_to_growth.html (дата обращения 31.01.2015).
9. The Montreal protocol on substance that depletes the ozone layer. United Nations Environment Programme, Ozone Secretariat, 1989 [Электронный ресурс]: http://ozone.unep.org/new_site/en/montreal_protocol.php (дата обращения 31.01.2015).
10. Kyoto Protocol. United Nations Framework Convention on Climate Change, 1992 [Электронный ресурс]: Available at: http://unfccc.int/2860.php (дата обращения 31.01.2015).
11. The history of sustainable development in the United Nations // United Nations Conference on Sustainable Development, 2012 [Электронный ресурс]: http://www.uncsd2012.org/history.html (дата обращения 31.01.2015).
12. United Nations Millennium Declaration. United Nations, 2000 [Электронный ресурс]: http://www.un.org/millennium/declaration/ares552e.htm (дата обращения 31.01.2015).
13. Document WSIS-03/GENEVA/DOC/4-E, 12 December 2003 // World Summit on the Information Society, 2003 [Электронный ресурс]: http://www.itu.int/net/wsis/docs/geneva/official/dop.html (дата обращения 31.01.2015).
14. Report of the Working Group on Internet Governance, Chateau de Bossey, June 2005 // Working Group on Internet Governance, 2005 [Электронный ресурс]: http://www.wgig.org/docs/WGIGREPORT.pdf (дата обращения 31.01.2015).
15. United Nations General Assembly Resolution A/Res/67/195. United Nations, 2012) [Электронный ресурс]: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/67/195 (дата обращения 31.01.2015).
16. The mapping of international Internet public policy issues. Conference room paper. Geneva: United Nations CSTD Secretariat, 2015. 22 pp.

SPECIFICS OF PROJECTS IN THE AREA OF ENTERPRISE ARCHITECTURE DEVELOPMENT

Dmitrij V. KOZNOV
Associate Professor, Department of Software Engineering, Saint Petersburg State University
Address: 28, Universitetsky Prospect, Stary Peterhof, St. Petersburg, 198504, Russian Federation
E-mail: st008149@spbu.ru

Maxim Yu. ARZUMANYAN
Assistant Professor, Department of Information Technologies in Economics,
The Bonch-Bruevich Saint Petersburg State University of Telecommunications;
Smart Architects Ltd.
Address: 22, Prospect Bolshevikov, St. Petersburg, 193232, Russian Federation
E-mail: maxim.arzumanyan@gmail.com

Yury V. ORLOV
General Director, Smart Architects Ltd.
Address: 61A, Moyka, St. Petersburg, 191186, Russian Federation
E-mail: yury.orlov@smartarchitects.ru

Maxim A. DEREVYANKO
Engineer, Department of Information Technologies in Economics, The Bonch-Bruevich
Saint Petersburg State University of Telecommunications; Smart Architects Ltd.
Address: 22, Prospect Bolshevikov, St. Petersburg, 193232, Russian Federation
E-mail: maxim.derevyanko@gmail.com

Konstantin Yu. ROMANOVSKY
Senior Lecturer, Department of Software Engineering, Saint Petersburg State University
Address: 28, Universitetsky Prospect, Stary Peterhof, St. Petersburg, 198504, Russian Federation
E-mail: k.romanovsky@spbu.ru

Anastasia A. SIDORINA
Student, Department of Computer Science, Saint Petersburg State University
Address: 28, Universitetsky Prospect, Stary Peterhof, St. Petersburg, 198504, Russian Federation
E-mail: sidorina-nastia@mail.ru

Enterprise Architecture (EA) is a discipline for business and IT system management, describing the fundamental artifacts of business and IT. Introducing EA activities in a company is done via special projects (referred to as EAM projects). While much research focuses on EA as is, EAM projects are not yet considered properly. This paper deals with EAM projects as a special kind of IT project. The definition of an EAM project is suggested using the IT solution concept of the Microsoft Solution Framework. This is justified, since the EAM-software introduced in a company is the main result of a typical EAM project. Based on the definition of a work product in CMMI, this paper introduces the final delivery of an EAM project. In addition, domain specific modeling is used to describe EAM project delivery, since the most important part of EA is concepts and terms for describing business and IT in a company. To implement the language formed by these terms and concepts, an EAM tool is selected and customized (Mega, IBM System Architects, Aris, etc.), and additional software (i.e. Web-portal, integration scripts, etc.) is implemented. All of the EA principles and methods could be considered as guidelines for this software (EAM tool and additional software). Final delivery of an EAM project is divided into the following parts: method (domain specific language, EA method, EA process, integration, modeling results), technology (EAM tool, additional software, documentation), support and training. Using the concepts introduced in this paper, four EAM projects are analyzed. Lack of support of the EA method is identified as a common problem of these projects.

Introduction

The modern market is mobile and volatile due to a high competition, changing technology, as well as external circumstances (political, geographical and so on.). Therefore, companies to be successful, have to change on a permanent basis. To ensure efficient management and renovation of companies, the scientific and practical area referred to as *Enterprise Architecture Management* (EAM) – analysis, design, planning and implementation of the company activity based on a systemic concept [7, 21] is currently rapidly growing. This activity is supported by special software products (hereinafter referred to as EAM-tools), such as ARIS [9], Mega [22] and some others. The introduction of architecture management progress is generally accompanied by creation of solutions based on EAM-tools, and in most cases implemented within special projects, which will be called EAM-projects. In implementation of EAM-projects, standard tools are adapted to the individual peculiarities of the customer company, namely methodologies and standards (TOGAF, DoDAF, Archimate et al. [12]), graphic languages and notations [16, 19, 20], as well as software tools [11].

The purpose of this paper is to define EAM-projects, considering them as specific IT projects. The resulting delivery of an EAM-project is also investigated, with the identification of its typical Work Products [13] that will make it possible to define the scope of implementation and scope of works of such projects. In addition to the above, methods of Software Engineering and Domain-Specific Modeling (DSM) are used [18].

1. Basic definitions

1.1. Enterprise Architecture Management

Enterprise Architecture[1] (EA) covers the basic principles of its existence and its development either as an isolated structure, or together with partners, suppliers and/or buyers. Specifically, the enterprise can be considered as a whole or partially (for example, any single business area or department) [24]. Enterprise architecture is generally divided into IT architecture and business architecture. The development of IT architecture involves systematizing and coordinated development of the company's IT infrastructure, i.e. information systems, platforms, networks, hardware, etc. The business architecture management consists in identification and description of the business tasks of the company, its functional-role structure, business processes, etc.

Visual modeling plays the key role in the implementation projects of architecture management: the company activity and infrastructure are described by a set of models that makes these descriptions available to a wide range of specialists. The rapidly growing market of EAM-tools [11] is largely a modeling tools market.

1.2. CMMI and Work Products

In the early 1990s, the CMM (Capability Maturity Model) standard for certification of military and federal contractors in the area of software development was developed in the United States. In 2000, a generalized standard version called CMMI (Capability Maturity Model Integrated) and combined development of software and non-software components of the complicated artificial systems was released. Currently this standard is widely used in global industry and, among other things, is a known glossary of IT terms. In particular, it provides a definition of a Work Product as a significant result of the development process which can be a software component, document, invoice, established process (for example, process of using software by end-users), etc. [13].

---

[1] The enterprise is taken to mean a business company or government agency.

1.3. IT solution in MSF

Early in 2000, Microsoft created a software development methodology called the Microsoft Solution Framework (MSF). One of the basic MSF concepts was the IT solution – software that is not «packaged» and is created to meet the needs of a particular customer. An IT solution in MSF is defined as the coordinated delivery of a set of elements (work products in CMMI terms) required to meet some business need of a specific customer. The examples can include an application software code, documentation, communications, implementation process, training, support [23].

1.4. DSM-approach
in software engineering

Domain specific modeling is a policy of creating software that increases the level of abstract development compared to traditional means by providing applications specifications in terms of the data domain [18].

This paper [5] draws parallels between software engineering and enterprise architecture management. In particular, it is shown that the use of visual modeling has much in common in these areas: it is a marketable product, and mature software tools are widely used in these areas. Finally, domain specific modeling is applied in these areas, thereby making it possible to configure standard tools (languages, methods, software modeling tools) for a specific task or develop new tools. This paper [5] also generalizes a definition of DSM solutions oriented in both data domains. However, it happened to be too general and did not cover all work products included in EAM-project delivery.

2. EAM-project

Generally, the company architecture development activities are organized in the form of one or several projects: a company is able to focus on the next round of changes and unification by providing appropriate resources. It is reasonable to create methods and tools to solve company problems within such projects. However, final implementation of the proposed methods and tools in the management processes, as well as their full use, shall be carried out by the company on its own. Therefore, it is important that the architecture description can be added and changed, and that this can be done by company employees. To do this, in the course of the EAM-project an appropriate IT solution is implemented in the company.

There are many projects (especially in Russia) on development company architecture which are restricted to generation of static descriptions of company management processes (as is or to be models). There are also projects within which new regulations, standards and formal procedures are generated. This can also be an EA-activity if it is implemented based on unified principles. To accomplish this, both activities are often performed by external consulting organizations. However, we will call EAM-projects only those architecture management projects within which EAM-tools are implemented (often with modification), and which, therefore, have a significant IT component.

3. Definition
of EAM-solution

Let us call the EAM project result an EAM-solution. Its main difference from MSF IT solution is that in the first case, the software percentage in the project is substantially lower than in the second. As a platform for implementation of EAM-solutions, ready modeling tools are used, with significant efforts spent to study the company work specifics, integrate solutions with various processes into companies, as well as develop documentation, coordinate and implement the solution.

Let us divide the EAM-solution into three parts, i.e. methodological unit, technology unit and maintenance and support unit. The methodological unit includes a modeling language, modeling procedure, process integration and results. The technological unit is composed of a basic tool and set of software tools and documentation. Finally, the maintenance and support unit comprises two elements for training and support. From the experience of our projects, we have concluded that such splitting is reasonable in organization and performance of EAM-projects: it seems to be more «operational» than various classifications of the composition of enterprise architecture [15, 17].

Let us turn our attention in more detail to the description of work products that are part of each unit.

A modeling language is created as part of the EAM-project, inasmuch as large companies generally have unique, specific characteristics, with the result that standard languages and notations proposed within EAM-tools need to be improved. In so doing, one uses extenders which are generally available for all EAM-tools [11]. Therefore, domain-specific modeling languages appear in this area [18]. In developing such a language, it is important to define the concepts that shall be used in modeling: with their help it is convenient for

company specialists to describe the enterprise activities. For this purpose, reference models, industry standards and frameworks are often used and, in fact, explicitly or implicitly ontologies are created [1, 2]. The modeling language can be a simple slice/setting of a standard language, but a more complicated case is possible, when the mathematical model of the base language is changed, and new concepts and interconnections, new modeling viewpoints, types of charts, etc. are defined.

In order to determine the proper use of the language modeling a *procedure* is created. The main question to which it gives an answer is «how to model» («what to model» was defined when creating the language). The procedure contains modeling scenarios, detail levels and modeling focuses in development of various models, as well as sources of information for development / modification of the models.

An important work product is a running and efficient architecture management process. Among the main aspects to be covered by this process are:

✦ roles of the solution users including various rights to change information in the repository;

✦ end-to-end scenarios of using solutions — both basic and system (support of the integrity and correctness of the information repository, backup, version control, etc.);

✦ additional rules, i.e. naming objects, storage folder for charts and objects, rules to work with predefined objects, etc.

Such a process is started.

Integration is implementation of the interrelations of the architecture management process with other processes and company functions (requirements, project management and so on.). There is also a need to implement the EAM-solution integration with various company information systems to exchange data with them during architecture management. Various techniques and standards, specifically, TOGAF [24], give great attention to this work product.

An important part of the final delivery covers *modeling results* of the company architecture. This part of the delivery can be created by the project team and/ or company specialists. Meanwhile, «as is» models, i.e. a description of the current situation of the company, and «to be» model, i.e. a description of the final (desired) organization of the company are created. The creation of these models is a time-consuming activity, so you need to clearly define exactly what kind of models should be developed within the project and their details.

Selection of an appropriate *basic tool* is an important work product of the EAM-project. The EAM functionality is described many times and in detail (see, for example [11]). Many organizations, for example, Gartner [11] are involved in EAM-tools classification. These approaches are used by the majority of the business enterprises when selecting EAM-tools. However, the selection of the EAM-tool can in addition be affected by the following circumstances:

✧ if your company is already using a certain EAM-tool, then, it is reasonable to use it in the EAM-project;

✧ when selecting a foreign EAM-tool for Russian companies, it is important that the supplier has a mission/partner in the Russian Federation to provide support for the implementation and operation of the tool; otherwise, the use of such a tool has high risks.

The software environment is the most important set of work products of the EAM-solution. This environment is created by configuring/modification of the basic EAM-base tool. *Configuring* involves the use of EAM-tool settings (templates, palettes, etc.). *Modification* is the creation of additional software using an open software EAM-tool interface. Let us select the following work products of the software environment:

✦ *user interfaces* intended for different user groups;

✦ *basic language support*, allowing users to work in EAM-tool in terms of the created modeling language;

✦ *integrity support* (*post-processing*) — packet auditors for diagram syntax (basically required, as the modeling language is usually hard to implement in the EAM-tool);

✦ *integrity support* (*administration*) — monitoring the correct location of individual model elements in the repository packs, rules of using predefined objects, cleaning the model from «garbage», etc.;

✦ *import/export of data* — data exchange between the EAM-tool and different information systems existing in the company;

✦ model *reports generators* — configuring the existing reports and creating new ones;

✦ *web-portal* — contains reference data on the modeling, often the modeling results. Although EAM-tools often support automatic generation of portals for the repository, many projects require the creation of a complicated and multi-functional portal (see the example in paper [10]);

✦ *repository structure (*in terms of ARIS [9] — the folders in terms of [6, 8] — classifiers); when designing the

structure, not only the information structure, but also the areas of responsibility in the company are taken into account (that is, the repository shall have no folders, for which nobody is responsible);

✦ *scanning and setting up* the created EAM-solution.

Let us note that the methods and volumes of implementation of these work products are dependent on the properties of the basic EAM-tool used in the project.

All basic delivery elements must be provided with *documentation*. The main objective of the documentation is to facilitate as much as possible the involvement of company employees/contractors in the architecture management processes. The documentation may also be posted on the web-portal, contain a significant amount of visual materials, including video tutorials, quizzes to test knowledge, and more. The volume and complexity of this part of delivery can vary profoundly, becoming one of the most resource-intensive parts of the EAM-solution.

An essential EAM-solution component is *training*. The result of training is that the customer is able to use the solution all by himself.

Table 1.

Methodological unit

#	Final delivery elements	EAM projects			
		P1	P2	P3	P4
1	Modeling language	+ + +	+	+ +	+ +
2	Procedure	-	-	-	+ -
3	Process	+	+ -	+	+ -
4	Integration	+ -	-	+	-
5	Modeling results: «as is»	+ +	+	+	+
	Modeling results: «to be»	-	+ -	+	-

Solution *maintenance* means assistance to users in working with the solution upon completion of the project, as well as correction of a number of errors and shortcomings. The scope of work that can be performed within the maintenance is normally stipulated in the contract.

Table 2.

Technological unit

#	Final delivery elements	EAM projects			
		P1	P2	P3	P4
1	Basic EAM-tool	ARIS	Mega	IBM System Architect	ORG-Master (Russian EAM-package)
2	Software environment				
2.1	User interfaces	+ -	+ -	+	+
2.2	Basic support of the modeling language	+ -	+	+	+ -
2.3	Integrity support (online)	+ -	+	+	+ -
2.4	Integrity support (post-processing)	+	+	+	-
2.5	Integrity support (administration)	+	+	+	+ -
2.6	Import/export	+	+	+	-
2.7	Web-portal	+	+ +	+	+ +
2.8	Report generators	+ -	+ -	+	+ +
2.9	Repository structure	+	+	+	+
2.10	Solution scanning and setting up	+	+	+	+
3	Documentation	+	+	+	+ -

4. Examples

Tables 1, 2 and 3 provide descriptions of EAM-projects in terms of the EAM-solution. Projects P1, P2, P3 were performed for large state-owned corporations of the Russian Federation (IT-architecture development), while project P4 was implemented for a large government organization of the Russian Federation (business architecture development). Of all the projects, only P2 was not completed successfully, although again its results (ontology and modeling language) were used by the company. The most complicated domain-specific language was created within project P1: it was developed almost «from scratch» and used the product language ARIS as an implementer. However, this language maintenance was inferior to similar maintenance in projects P2 and P3 (items 2.2 and 2.3 in *Table 2*). This resulted from the fact that the ARIS packet is considerably inferior to Mega and IBM System Architect in terms of functional capabilities of the extenders. The EAM-packet used in project P4 did not support visual modeling tools. Therefore, not only a domain-specific modeling language but an ontology (i.e., without diagrammatic notation) was created within the project, with the result that the integrity support information in this ontology was negligible. The user interface in project P3 was created based on Share Point and turned out to be the most multi-purpose of all projects (item 2.1 of *Table 2*).

Conclusion

This paper addresses a definition of the EAM-project and its result — an EAM-solution. Review of the EAM-project as a specific IT project made it possible to apply software engineering tools to its description. A precise identification of components of the final delivery of EAM-solution can help in devel-

Table 3.

Maintenance and support unit

#	Final delivery elements	EAM projects			
		P1	P2	P3	P4
1	Training	+	-	+	+ -
2	Maintenance	+	-	+	+ -

opment of requirements in EAM-projects and evaluation of resources necessary for their implementation. The model can be also used to assess the quality of EAM-project results. In addition, this model is applicable for further study of EAM-projects. Even a fairly superficial analysis of the four projects carried out with its help has revealed the following tendency, i.e. weak support of the modeling technique: specifically, customers do not understand the importance of this part of delivery combining the technique with the user documentation or even excluding it from the final delivery.

For a more efficient practical application of the EAM-solution final delivery model, numerical metrics should be created to evaluate its work products. Further work is also needed to deepen the parallels between the domain specific modeling in software engineering and architecture management. Generally speaking, the results of this work are a first step in application of software engineering techniques to studying projects in the area of enterprise architecture.

Furthermore, an important aspect of a more successful implementation of EAM-projects is the issue of high-quality training of company employees (within or beyond the framework of the EAM-project), the latest architecture management method, knowledge management, etc. [3, 4]. ■

References

1. Gavrilova T.A. (2002) Logiko-lingvisticheskoe upravlenie kak vvedenie v upravlenie znanijami [Logic-linguistic methods as introduction in knowledge management]. *Novosti iskusstvennogo intellekta*, no. 6, pp. 22—27 (in Russian).

2. Gavrilova T.A., Kudrjavcev D.V., Gorovoj V.A. (2006) Modeli i metody formirovanija ontologij [Models and methods for ontology development]. *Nauchno-Tehnicheskie Vedomosti SPbGPU*, no. 46, pp. 21—28 (in Russian).

3. Gavrilova T.A., Leshcheva I.A., Kudrjavcev D.V. (2012) Ispol'zovanie modelej inzhenerii znanij dlja podgotovki specialistov v oblasti informacionnyh tehnologij [Using of knowledge engineering models for training of specialists in the field of information technologies]. *System Programming*, vol. 7, no. 1, pp. 90—105 (in Russian).

4. Gavrilova T.A., Leshcheva I.A., Strahovich E.V. (2011) Ob ispol'zovanii vizual'nyh modelej v prepodavanii [On application of visual models in teaching]. *Herald of Saint Petersburg University. Series 8: Management*, no. 4, pp. 124—150 (in Russian).

5. Koznov D.V. (2014) *Vizual'noe modelirovanie informacionnyh e-servisov v publichnoj sfere* [Visual modeling in development of public information e-services]. St.Petersburg: SPbGU (in Russian).

6. Grigor'ev L.Ju., Koznov D.V., Kudrjavcev D.V. (2013) Obzor jazyka modelirovanija ORG-Master [OGR-Master modeling language]. *Sistemnoe programmirovanie*, vol. 8, no. 1, p. 5 (in Russian).

7. Kudrjavcev D.V., Arzumanjan M.Ju., Grigor'ev L.Ju. (2014) *Tehnologii biznes-inzhiniringa* [Business engineering technologies]. St.Petersburg: SPbGPU (in Russian).

8. Kudrjavcev D.V., Koznov D.V., Grigor'ev L.Ju. (2013) Sredstva tipizacii i pragmatika jazyka modelirovanija ORG-Master [Typing and pragmatics of ORG-Master modeling language]. *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki*, vol. 88, no. 6, pp. 79—85 (in Russian).

9. Kamennova M., Gromov A., Ferapontov M. (2001) *Modelirovanie biznesa. Metodologija ARIS. Prakticheskoe rukovodstvo* [Business modeling: ARIS guidelines]. Moscow. Vest' — Meta Tehnologija (in Russian).

10. Kostyrko A., Kudrjavcev D., Grigor'ev L., Kislova V., Zhulin A., Sinyatullina L., Ermakov R. (2012) Modelirovanie kompleksov gorodskogo hozjajstva dlja sistemnogo razvitija IKT goroda [Modeling of municipal facilities for system development of municipal info-communication technologies]. Proceedings of the *Knowledge Engineering and Semantic Web Technologies Conference (KESW 2012), November 2-8, 2012, Saint Petersburg*, pp. 81—88 (in Russian).

11. Brand S. (2014) *Magic Quadrant for Enterprise Architecture Tools*. Gartner.

12. Buckl S., Schweda C.M. (2011) *On the state-of-the-art in Enterprise Architecture Management literature: Technical report*. Munchen: Technische Universitat Munchen.

13. *CMMI-SVC, CMMI for Services*, Version 1.3. (2010) Available at: http://www.sei.cmu.edu/reports/10tr034.pdf (accessed 27 January 2015).

14. Cox I.R. (2014) *Enterprise Architecture: How to get EA optimized*. Amazon Digital Services.

15. Federal CIO Council (2013) *Federal Enterprise Architecture Framework*. Version 2. Federal CIO Council, US.

16. Frank U. (2014) Enterprise modelling: The next steps. *Enterprise Modelling and Information Systems Architectures*, vol. 9, no. 1, pp. 22—37.

17. ISO (2000) *ISO 15704. Industrial automation systems — Requirements for enterprise-reference architectures and methodologies*. ISO.

18. Kelly S., Tolvanen J. (2008) *Domain-specific modeling: Enabling full code generation*. Wiley, IEEE Computer Society Press.

19. Kudryavtsev D., Gavrilova T. (2011) Diagrammatic knowledge modeling for managers: Ontology-based approach. Proceedings of the *International Conference on Knowledge Engineering and Ontology Development, (KEOD 2011), October 26, 2011, Paris, France*, pp. 386—389.

20. Kuhrmann M., Kalus G., Knapp A. (2013) Rapid prototyping for domain-specific languages — From stakeholder analyses to modelling Tools. *Enterprise Modelling and Information Systems Architectures*, vol. 8, no. 1, pp. 62—74.

21. Lankhorst M.M. (2005) *Enterprise Architecture at work: Modelling, communication and analysis*. Berlin, Heidelberg: Springer.

22. MEGA International (2011) *MEGA Studio User Guide*. 2nd Edition. MEGA International.

23. *Microsoft Solutions Framework, Process Model*, Version 3.1. (2002) Microsoft. Available at: http://www.microsoft.com/msf (accessed 28 January 2015).

24. TOGAF, Version 9.1. Enterprise Edition. (2011) Available at: http://www.opengroup.org/togaf/ (accessed 25 January 2015).

25. Franke U., Ekstedt M., Lagerstrom R., et al. (2010) Trends in Enterprise Architecture practice — A survey. Proceedings of *5th International Workshop (TEAR 2010), November 12, 2010, Delft, The Netherlands*, pp. 16—29.

ОСОБЕННОСТИ ПРОЕКТОВ В ОБЛАСТИ РАЗРАБОТКИ КОРПОРАТИВНОЙ АРХИТЕКТУРЫ ПРЕДПРИЯТИЙ

Д.В. КОЗНОВ
кандидат физико-математических наук, доцент кафедры системного программирования, Санкт-Петербургский государственный университет
Адрес: 198504, г. Санкт-Петербург, Старый Петергоф, Университетский проспект, д. 28
E-mail: st008149@spbu.ru

М.Ю. АРЗУМАНЯН
ассистент кафедры информационных технологий в экономике, Санкт-Петербургский государственный университет телекоммуникаций им. М.А.Бонч-Бруевича; ООО «Смарт Архитектс»
Адрес: 193232, Санкт-Петербург, проспект Большевиков д. 22, корп. 1
E-mail: maxim.arzumanyan@gmail.com

Ю.В. ОРЛОВ
генеральный директор ООО «Смарт Архитектс»
Адрес: 191186, г. Санкт-Петербург, наб. реки Мойки, 61А
E-mail: yury.orlov@smartarchitects.ru

М.А. ДЕРЕВЯНКО
инженер кафедры информационных технологий в экономике, Санкт-Петербургский государственный университет телекоммуникаций им. М.А.Бонч-Бруевича; ООО «Смарт Архитектс»
Адрес: 193232, Санкт-Петербург, проспект Большевиков д. 22, корп. 1
E-mail: maxim.derevyanko@gmail.com

К.Ю. РОМАНОВСКИЙ
кандидат физико-математических наук, старший преподаватель кафедры системного программирования, Санкт-Петербургский государственный университет
Адрес: 198504, г. Санкт-Петербург, Старый Петергоф, Университетский проспект, д. 28
E-mail: k.romanovsky@spbu.ru

А.А. СИДОРИНА
студентка кафедры информатики, Санкт-Петербургский государственный университет
Адрес: 198504, г. Санкт-Петербург, Старый Петергоф, Университетский проспект, д. 28
E-mail: sidorina-nastia@mail.ru

Управление архитектурой предприятия (Enterprise Architecture Management, EAM) является ключевым видом деятельности при повышении эффективности бизнеса в современных компаниях. Данная деятельность поддерживается специальными программными продуктами (так называемыми EAM-инструментами) — ARIS, Mega и др. Внедрение процесса управления архитектурой, как правило, происходит путем реализации соответствующих проектов. Данная статья посвящена исследованию таких проектов. Среди них выделяются так называемые EAM-проекты, которые являются специфическими ИТ-проектами и направлены на создание и развертку в компании ИТ-решения по управлению архитектурой на основе какого-либо EAM-инструмента. Для дальнейшего исследования EAM-проектов использованы методы программной инженерии (Software Engineering) — методология разработки ПО Microsoft Solution Framework (MSF) и терминология стандарта CMMI, — а также концепция предметно-ориентированного моделирования (Domain-Specific Modeling, DSM).

В работе детально описана финальная поставка типового EAM-проекта, подразделяющаяся на методологический блок (язык моделирования, методика, процесс, интеграция, результаты моделирования), технологический блок (базовый инструмент, комплекс программных средств, документация), сопровождение и поддержку (обучение, сопровождение). Обсуждаются особенности разработки этих рабочих продуктов. При этом детально анализируется программная часть поставки — пользовательские интерфейсы, базовая поддержка языка моделирования, поддержка целостности (online), поддержка целостности (post-processing), поддержка целостности (administration), импорт/экспорт, Web-портал, генераторы отчетов, структура репозитория, развертка и настройка решения. В рамках введенных понятий обсуждаются четыре EAM-проекта по реализации управления архитектурой, реализованные в российских компаниях: три проекта были выполнены для различных госкорпораций, один — для крупного органа государственной власти. Выявлена типовая сложность этих проектов — недостаточная поддержка методики моделирования.

Литература

1. Гаврилова Т.А. Логико-лингвистическое управление как введение в управление знаниями // Новости искусственного интеллекта. 2002. № 6. С. 22—27.
2. Гаврилова Т.А., Кудрявцев Д.В., Горовой В.А. Модели и методы формирования онтологий // Научно-технические ведомости СПбГПУ. 2006. № 6. С. 21—28.
3. Гаврилова Т.А., Лещева И.А., Кудрявцев Д.В. Использование моделей инженерии знаний для подготовки специалистов в области информационных технологий. Системное программирование. 2012. Т. 7, № 1. С. 90—105.
4. Гаврилова Т.А., Лещева И.А., Страхович Э.В. Об использовании визуальных моделей в преподавании. Вестник Санкт-Петербургского университета. Серия 8: Менеджмент. 2011. № 4. С. 124—150.
5. Кознов Д.В. Визуальное моделирование информационных е-сервисов в публичной сфере. СПб.: СПбГУ, 2014. 144 с.
6. Григорьев Л.Ю., Кознов Д.В., Кудрявцев Д.В. Обзор языка моделирования ОРГ-Мастер // Системное программирование. 2013. Т. 8, № 1. С. 5.
7. Кудрявцев Д.В., Арзуманян М.Ю., Григорьев Л.Ю. Технологии бизнес-инжиниринга. СПб.: СПбГПУ, 2014. 427 с.
8. Кудрявцев Д.В., Кознов Д.В., Григорьев Л.Ю. Средства типизации и прагматика языка моделирования ОРГ-Мастер // Научно-технический вестник информационных технологий, механики и оптики. 2013. Т. 88, № 6. С. 79—85.
9. Моделирование бизнеса. Методология ARIS. Практическое руководство / М.С. Каменнова и [др.]. М.: Весть — Мета Технология, 2001. 333 с.
10. Моделирование комплексов городского хозяйства для системного развития ИКТ города / А. Костырко и [др.] // Сб. трудов конференции «Инженерия знаний и технологии семантического веба» (KESW 2012), 2-8 ноября 2012 г., г. Санкт-Петербург. С. 81—88.
11. Brand S. Magic Quadrant for Enterprise Architecture Tools. G00262. Gartner, 2014. 38 p.
12. Buckl S., Schweda C.M. On the state-of-the-art in Enterprise Architecture Management literature: Technical report. Munchen: Technische Universitat Munchen, 2011. 118 p.
13. CMMI-SVC, CMMI for Services, Version 1.3. 2010. 520 p.
14. Cox I.R. Enterprise Architecture: How to get EA optimized. Amazon Digital Services, 2014. 48 p.
15. FEAF-II. Federal Enterprise Architecture Framework. Version 2. Federal CIO Council, US. 2013. 434 p.
16. Frank U. Enterprise modelling: The next steps // Enterprise Modelling and Information Systems Architectures. 2014. Vol. 9, No. 1. P. 22—37.
17. ISO 15704. Industrial automation systems — Requirements for enterprise-reference architectures and methodologies. 2000. 7 p.
18. Kelly S., Tolvanen J. Domain-specific modeling: Enabling full code generation. Wiley, IEEE Computer Society Press, 2008. 444 p.
19. Kudryavtsev D., Gavrilova T. Diagrammatic knowledge modeling for managers: Ontology-based approach // Proceedings of the International Conference on Knowledge Engineering and Ontology Development (KEOD 2011), October 26, 2011, Paris, France. P. 386—389.
20. Kuhrmann M., Kalus G., Knapp A. Rapid prototyping for domain-specific languages — From stakeholder analyses to modelling tools // Enterprise Modelling and Information Systems Architectures. 2013. Vol. 8, No. 1. P. 62—74.
21. Lankhorst M.M. Enterprise Architecture at work: Modelling, communication and analysis. Berlin, Heidelberg: Springer, 2005. 338 p.
22. MEGA Studio User Guide. 2nd Edition. MEGA International, 2011. 114 p.
23. Microsoft Solutions Framework, Process Model, Version 3.1 // Microsoft 2002. [Электронный ресурс]: http://www.microsoft.com/msf (дата обращения 28.01.2015).
24. TOGAF, Version 9.1. Enterprise Edition. 2011. 692 p.
25. Trends in Enterprise Architecture practice — A survey / U. Franke [et al.] // Trends in Enterprise Architecture Research. Proceedings of 5th International Workshop (TEAR 2010), November 12, 2010, Delft, The Netherlands. P. 16—29.

MODEL OF AN OPTIMAL PUBLIC-PRIVATE PARTNERSHIP PROJECT IN THE TELECOMMUNICATIONS SECTOR

Svetlana V. MALTSEVA

Professor, Department of Innovation and Business in Information Technologies,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: smaltseva@hse.ru

Polina V. KOTELNIKOVA

Post-graduate Student, Department of Innovation and Business in Information Technologies,
National Research University Higher School of Economics;
Chief Inspector, Russian Accounts Chamber
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: kotelnikovapolina@gmail.com

This article presents an investment project model in the telecommunications sector using public-private partnership (PPP), making it possible to link the main project parameters (duration, volume of investments, fare, request) with the expected indexes of effectiveness. When considering the project parameters and indexes of effectiveness, interests of both the state and private company are taken into account.

The modeling algorithm and criteria of performance evaluation are developed based on the standards approved for evaluation of investment projects with state participation, namely the Guidelines on performance evaluation of the investment projects approved by the Ministry of Economy, Ministry of Finance, State Construction Committee of the Russian Federation on June 21, 1999, No. VK 477.

This paper assumes that the most important criterion for evaluation of project effectiveness of a private company is the maximum of the net present value of the project, meaning the excess of the total cash receipts over the total cost for the project with regard to disparity of the effects (costs and benefits) related to different moments in time. Therefore, the higher this index, the greater is the interest of a private company to participate in the project.

It seems that from the standpoint of the state two factors are most important: the social significance of the project and minimum expenses of the state for implementation of the investment project under conditions of a limited capacity of spending budget funds. The social significance of the project is defined by expert study as the impact of the operational results on at least one of the domestic or foreign markets: financial markets, product and service markets, labor market, etc., as well as the ecological and social environment.

The model makes it possible to calculate various scenarios to determine optimal project parameters ensuring maximum efficiency for a given limit on the amount of budget investments.

The calculation results for the proposed model can be used to make decisions to participate in the project by the authorities or public development finance institutions. It is proposed to document the model used and its evaluation criteria in the rules of granting subsidies for such projects.

Introduction

The telecommunications sector is a high-tech industry and plays an important role in the development of the national economy and its competitiveness in response to global challenges. Its development not only creates better conditions for the national business activity, but also a favorable climate to attract foreign investments in other sectors of the national economy, because a good infrastructure development level is a prerequisite for effective and efficient business operations. Under conditions of shortage and high prices on the national market of investment resources, attracting state support is a high priority for Russian enterprises in the telecommunications sector.

In big cities, the telecommunications market is nearly saturated, and significant growth in the number of subscribers is possible only by development of new, most often hardly accessible areas where they live. At the same time, the basic limitation for operators in implementation of projects aimed at attracting new subscribers in hardly accessible regions is the need for large-scale investments in basic telecommunications infrastructure in these regions. However, through introduction of operators' innovations in the internal business processes, the cost maintenance value of a new subscriber will be low. Therefore, when implementing such infrastructure projects, operators rely on state support.

At the same time, one of the key objectives of the state in the area of telecommunications is to ensure that citizens have access to modern telecommunications services irrespective of their place of residence. This is stated in the government program of the Russian Federation «Information Society (years 2011-2020)» [4]. This program envisages establishment of a modern information and telecommunications infrastructure across Russia and reduction of the «digital gap» of some federal territorial unites of the Russian Federation.

The issue of public-private partnership (PPP) in the telecommunications sector is also highly topical in world practice. The World Bank, in particular, pays close attention to PPP projects in the telecommunications sector. The official website of the World Bank [7] constantly updates information on the regulatory environment of the telecommunications sector in various countries (licensing, tariff regulation, universal communication services) and posts examples of implemented PPP projects (cases) in the European Union, Canada, USA, Australia and Singapore.

Total capital investments in the telecommunications sector in Russia in 2013 came to 270.7 billion rubles [4] and the commitment rate of the federal budget to solve government tasks in the telecommunications sector is no more than 15 billion rubles per year [1].

It is obvious that the state's capabilities to finance the telecommunications sector are much lower than the financial possibilities of the private sector. Therefore, the task of finding optimal PPP models between the state and the largest communication operators for development of state-of-the-art communications services in hardly accessible regions is extremely crucial.

The specifics of an investment project implemented under the terms and conditions of PPP are that it is important to calculate the efficiency of its implementation for the company at minimum cost of the state and on the condition that the project is socially minded. Therefore, when considering project implementation under PPP terms both the state and private company face the task of building a project model which will simultaneously take into account the interests of private companies and of society as a whole.

1. Algorithm for applying simulation modeling of the investment project in the telecommunications sector and its parameters

The implementation of investment projects in the area of cellular communication takes place under conditions of uncertainty, so even a qualitative business plan of the project cannot guarantee that under conditions of Russia's high-risk economy the investment project implemented can ensure the efficiency and profitability laid down in the plan.

Under conditions of uncertainty, one of the most reasonable approaches to the analysis and evaluation of investment projects is simulation modeling. Application of the modeling simulation in the efficiency analysis of investment projects makes it possible not only to make a reasonable decision about implementation or rejection of one or another investment project, but also identify scenarios leading to the best and the worst results of the project being analyzed. This, in turn, can contribute to the timely adjustment of the project implementation parameters in order to achieve maximum benefit on the investments and justification for applying the public-private partnership mechanisms.

For that reason, it becomes urgent to develop an integrated algorithm for applying simulation modeling in an investment analysis when working in the cellular communication market [5].

The following algorithm is proposed for applying simulation modeling to investment analysis in the telecommunications sector. It consists of eight stages (*Fig. 1*).

Phase 1. Formulation of the investment project analysis problems

| Acquisition of data about the investment project | Conceptual description of the investment project | Study of problem situation, research goal and task setting |

Phase 2. Creation of a conceptual model describing implementation specifics

Phase 3. Formalization of conceptual model and development of mathematical model of cash flows

| Translation of conceptual description elements into language of mathematical formula | Mathematical description of performance criteria of the investment project | Computation of criteria values of the project's effectiveness for the basic option of its implementation |

No ← Information is sufficient

Yes

Phase 4. Creation of simulation model to calculate the project's effectiveness criteria

| Selection of language and modeling environment | Development of the program-simulator |

Phase 5. Acquisition and preparation of data on incoming variables and models parameters

Phase 6. Validation and verification of the simulation model

| Evaluation of model suitability | Model verification | Data validity |

Suitability is confirmed → No

Yes

Phase 7. Performing the simulation experiment

| Generation of values of incoming variables | Generation of values of incoming parameters | Calculation of performance criteria for each simulation |

Phase 8. Analyzing the results

Fig. 1. General algorithm for applying simulation modeling to the investment analysis

The most important and critical stages of the proposed algorithm are the stages of constructing conceptual, mathematical and simulation models of the performance criteria of the investment project under consideration.

In the phase of conceptual model development the real system transfers to the logical scheme of its operation and the following steps are performed:

◇ determine the model structure – static and dynamic description of the investment project;

◇ define the system boundaries, describe the environment, as well as its most essential elements and properties which can affect the final result of the investment project under consideration;

◇ develop a list of random variables and deterministic parameters, functional dependencies, constraints and criteria of efficiency of the cellular operator's investment project. The variables and parameters should be selected from the following list of values that characterize the activities of the cellular communication operator: number of subscribers, call duration for different types of calls (within the operator's network, to numbers of other operators, to numbers of fixed-line telephony), the number of sms- and mms-messages sent, the amount of data transferred, the scope of other services provided, cost and net cost of one minute of a telephone call for various types of calls, cost and net cost of sending one sms- and mms-message, cost and net cost of transmitting one unit of data, cost and net cost of providing a unit of other services, value of dealership fees and other values at the discretion of the developers.

It is proposed to use the following values in the proposed model as parameters (constant values):

✦ amount of investments required to develop the telecommunications infrastructure;

✦ cost of one minute of outgoing call;

✦ cost of an outgoing message;

✦ cost of transmitting a unit of data;

✦ discount rate;

✦ project implementation period (period of construction and operation).

The project implementation period includes the period of facility construction and operation. Under standard conditions, the time required for completion of the construction cannot be precisely identified, since delays to the delivery of works is a common practice. On the other hand, if we do not specify the period of construction and specify only the project implementation period, the investor will be interested in reducing the construction period in order to maximize the facility operation period during which the investor derives revenue. This will have a negative impact on the quality of construction work. Therefore, the time period for facility construction and operation should be set as parameters of the investment project and indicated in the PPP contract.

The following values are proposed for assignment to the model variables:

✦ price of one minute of an outgoing call;

✦ price of an outgoing message;

✦ cost of transmitting a unit of data;

✦ number of subscribers;

✦ number of outgoing calls;

✦ average duration of an outgoing call;

✦ number of outgoing messages per subscriber;

✦ volume of data transmitted (traffic) per subscriber;

✦ government take in the investment financing.

For performing simulation experiment, the sample size must ensure its representativeness. This is determined for each random variable. At the end of the simulation experiment, a number of values of performance criteria of the investment project are expected to be received. For analysis of the modeling results, the law of distribution of the resulting indicator of the investment project is determined, and on this basis the project risks can be assessed. For this purpose, the expectation function and mean-square deviation indicators, as well as various probability measures, should be calculated. The result is that, for example, it will be fair to say with a definite probability that the value of the performance criterion of the investment project of the cellular operator will not be below a certain value, or determine the probability of obtaining values of performance criteria indicating the inefficiency of the investment project.

The application of the proposed algorithm will make it possible to quantify the risks of the investment project in the cellular market and make a justified decision about implementation of the investment project using the PPP mechanism, or decide against its implementation.

2. Criteria of effectiveness of the investment project in the telecommunications sector under the terms of a public-private partnership

The effectiveness of the investment project is a category reflecting the adherence of the project generating this investment project (IP) to the aims and interests of its participants [1].

It is common practice to refer the effectiveness of the project as a whole and the effectiveness of participating in the project to types of effectiveness of the investment projects (*Fig. 2*).
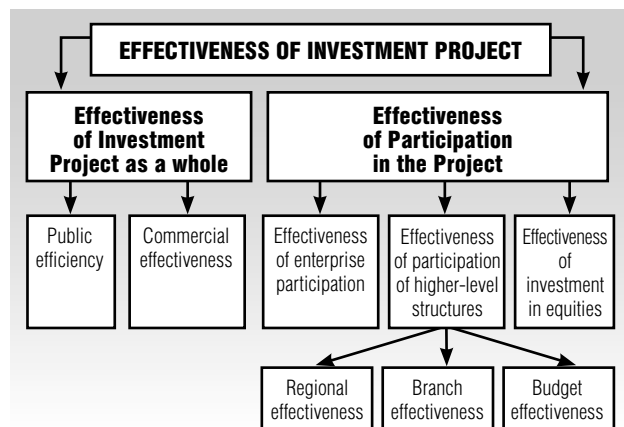
Fig. 2. Effectiveness of the investment project

The effectiveness of the project as a whole is estimated in order to determine the potential attractiveness of the project to potential participants and fundraising. It includes public (social and economic) efficiency and commercial effectiveness of the project.

The indicators of public efficiency account for social and economic consequences of the investment project implementation for society as a whole, including both immediate results and costs of the project, and «external» results, namely the costs and results in related economy sectors, environmental, social and other non-economic effects.

Indicators of the commercial effectiveness of the project account for financial implications of its implementation for the participant implementing the investment project, on the assumption that the participant bears all the necessary project costs and enjoy all its results.

The effectiveness of participation in the project is determined in order to verify the investment project implementation and interest of all project participants in the project.

Prior to assessment of effectiveness, the social significance of the project is determined by evaluation of experts. The social significance (scale) of the project is determined by the effect of the project implementation results on at least one of the internal or external markets: financial, products and services, labor and other markets, as well as on the ecological and social environment.

Further assessment is conducted in two phases.

The first phase calculates the effectiveness indicator of the project as a whole. For socially significant projects, their social efficiency is evaluated first. If the social efficiency is improper, such projects are not recommended for implementation and cannot qualify for state support. If their social efficiency turns out to be sufficient, their commercial effectiveness is assessed next. If the commercial effectiveness of the socially significant IP is insufficient, a recommendation is made to address the pos-

sibility of applying various forms of support which could increase the IP commercial effectiveness to an acceptance level. The subject of public-private partnership and selection of its implementation methods then comes out.

The second phase of assessment is performed after development of the funding model. This phase specifies a list of participants and identifies the financial feasibility and participation efficiency in each participant project.

On this basis, the following algorithm for constructing an optimal model of the PPP project is proposed in terms of its effectiveness (*Fig. 3*).
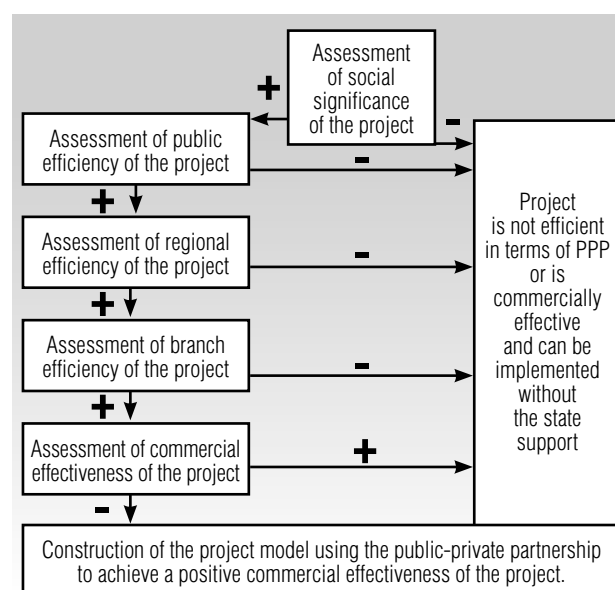
Fig. 3. Algorithm for constructing an optimal model of PPP project in terms of its effectiveness

It is recommended to use the following indexes as the main indicators to calculate IP effectiveness:

✦ net present value;
✦ internal rate of return;
✦ need for additional financing;
✦ profitability indexes of expenses and investments;
✦ payback period.

The conditions of financial feasibility and performance indexes are calculated on the basis of cash flow, whose specific components depend on the estimated kind of effectiveness.

The most important indicator of project effectiveness is **net present value** (NPV, another name is integral effect), i.e. the accumulated discounted effect for the accounting period. NPV is calculated by the formula:

$$NPV = \sum_m CF_m \frac{1}{\left(1+E\right)^{(t_m - t_0)}}$$

where CF_m is a cash flow at the end of m-th step;

E is the discount rate;

t_o is a beginning of a project;

t_m is an end of m-th step.

In order to recognize that the project if efficient in terms of the investor, the project *NPV* should be positive. When comparing alternative projects, priority shall be given to the project with a higher *NPV*, provided that *NPV* is positive.

Another effectiveness indicator is the **internal rate of return** (IRR, other names are internal rate of discount, internal rate of profit margin). In the most common case of IP beginning with investment outlay and having a positive net profit, the internal rate of return is a positive number E_0 which satisfies the following requirements:

❖ with a discount rate $E = E_0$ becomes zero in the net present value of the project;

❖ this is a singular number.

Payback period with regard to discounting is the period of time from the initial moment to payback moment with regard to discounting. The payback moment with regard to discounting is referred to as the earliest time in the accounting period after which the current net present value *NPV* (k) becomes greater than zero and keeps on negative. When assessing the effectiveness, the payback time usually acts as a limitation.

The need for additional financing with regard to discount (AF, other names are project costs, capital risk) is a maximum absolute value of the negative accumulated discounted balance of the investment and operational activities. The AF value indicates the minimum discounted amount of the external financing of the project required for its financial feasibility.

Profitability indexes characterize a (relative) project benefit per investment. They can be calculated both for discounted and non-discounted cash flows.

Net present value is proposed as an effectiveness criterion of investment projects in the telecommunications sector under public-private partnership conditions. It is proposed to find its maximum value with a given amount of investment and minimum state participation.

3. Investment project model for the telecommunications sector under conditions of public-private partnership: parameters and their association

The basic logic of the mathematic model construction procedure is to determine the parameters and variables that are included in the model, as well as the type of distribution, which these variables are subject to, and interdependences (functional and probabilistic relationships between variables).

Compliance with such a procedure is necessary to create a model that would be as follows:

$$NPV = f(x_1,..., x_i,..., x_n; a_1, ..., a_j, ..., a_m),$$

where x_i are variables (cash flow components being random values); n is a number variables;

a_j are fixed model parameters, i.e. those cash flow components which in the previous analysis were identified as independent or little dependent on the external environment, and therefore are considered as deterministic values;

m is a number of model parameters.

To construct a mathematical model of the investment project in the telecommunications sector under the terms of PPP, it is proposed to select the following model parameters and variables:

— model parameters:

✦ the amount of investments required for creation of telecom infrastructure (I);

✦ the cost of one minute of outgoing call of i-type in the t-th period (CX_{it});

✦ the cost of an outgoing message of i-type in the t-th period (CY_{it});

✦ the cost of transmission of a unit of data in the t-th period (CZ_t);

✦ fixed costs in the t-th period (FC_t);

✦ discount rate (r);

✦ duration of the project (T);

✦ period of project construction (T_i);

✦ useful lifetime of the project (T_0);

—model variables:

❖ price of a minute of outgoing call of i-type in the t-th period (PX_{it});

❖ price of an outgoing message of i-type in the t-th period (PY_{it});

❖ the cost of transmitting a data unit in the t-th period (PZ_t);

❖ number of outgoing calls of i-type per subscriber in the t-th period (X_{it});

❖ average duration of the i-type of outgoing calls for t-th period (D_{it});

❖ a number of outgoing messages of i-type per subscriber in the t-th period (Y_{it});

❖ amount of data transferred (traffic) per subscriber in the t-th period (Z_t);

❖ state participation in investment financing (j).

Therefore, the target function of the project can be written as follows:

$$NPV = \sum_{t=1}^{T_0 - T_i} \frac{N_t S - FC_t}{(1+r)^{T_0 - T_i}} - I(1-j),$$

where $S = P_{it}(PX_{it} - CX_{it}) + Y_{it}(PY_{it} - CY_{it}) + Z_t(PZ_t - CZ_t).$

The next investigation stage is calculation of the project function with specified project parameters and various options of possible values of variables with a restricted service prices set by a private company under the state anti-monopoly regulation and restricted capabilities of the public participation in the project.

Conclusion

The developed performance evaluation algorithm and criteria for the investment project on the terms of PPP made it possible to construct a model connecting the basic project parameters (duration, amount of investment, rate, demand) with the expected performance indexes.

The model identified seven parameters and nine variables.

The model makes it possible to calculate various scenarios to determine maximum effectiveness with restricted volume of state investments.

The results of calculation made on the basis of the proposed model can be used by the authorities or public financial institutions to reach a decision about participation in a project. It is proposed to document the model application and evaluation criteria in the rules of granting subsidies for the implementation of PPP projects.

The proposed algorithm for analysis of the investment projects under PPP terms has large-scale capabilities and makes it possible to adapt them to the available analysis models of investment projects, as well as modify them with due regard to actual conditions of implementation of the investment project.

The disadvantages of this approach are the subjectivity in selection of restrictions on the variables, including the volume of state investments. ∎

References

1. The Federal Law of 03.12.2012 No. 216-FZ (2012) O federal'nom bjudzhete na 2013 god i na planovyj period 2014 i 2015 godov [On the Federal Budget for 2013 and the planning period of 2014 and 2015 years]. *Rossiyskaya Gazeta*, no. 5956 (December 7, 2012), p. 19 (in Russian).

2. The Government of Russian Federation (2014) Gosudarstvennaja programma Rossijskoj Federacii «Informacionnoe obshhestvo (2011—2020 gody)» [State program of the Russian Federation «Information society (2011-2020)»]. *Collection of Russian Legislation*, no. 18 (May 5, 2014), p. 2159 (in Russian).

3. Kossov V.V., Livshits V.N., Shakhnazarov A.G. (2000) *Metodicheskie rekomendacii po ocenke jeffektivnosti investicionnyh proektov* [Guidelines for assessing the effectiveness of investment projects]. Moscow: Economics (in Russian).

4. The Ministry of Communications and Mass Communications of the Russian Federation (2013) *Investicii v otrasl' svjazi i vvod v dejstvie moshhnostej* [Investments in the communications industry and commissioning of facilities]. Available at: http://minsvyaz.ru/ru/pages/statistika-otrasli/#section-81 (accessed 25 March 2015) (in Russian).

5. Oblakova A. (2009) *Imitacionnoe modelirovanie investicion*noj dejatel'nosti v sfere uslug sotovoj svjazi [Simulation of investment activity in the field of mobile communication services]. PhD Thesis. Moscow, Finance Academy under the Government of the Russian Federation (in Russian).

6. Oblakova A. (2009) Imitacionnoe modelirovanie investicionnoj dejatel'nosti na rynke uslug sotovoj svjazi [Simulation of investment activity on the market of mobile communication services]. Proceedings of the *4th All-Russian Conference «Simulation. Theory and practice», 21-23 October, 2009, Saint Petersburg*, pp. 194—198 (in Russian).

7. The World Bank (2015) *Telecommunications / information & communication technology PPPs.* Available at: http://ppp.worldbank.org/public-private-partnership/sector/telecom (accessed 01 April 2015).

МОДЕЛЬ ОПТИМАЛЬНОГО ПРОЕКТА ГОСУДАРСТВЕННО-ЧАСТНОГО ПАРТНЕРСТВА В СЕКТОРЕ ТЕЛЕКОММУНИКАЦИЙ

С.В. МАЛЬЦЕВА
доктор технических наук, профессор кафедры инноваций и бизнеса в сфере информационных технологий, Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: smaltseva@hse.ru

П.В. КОТЕЛЬНИКОВА

аспирант кафедры инноваций и бизнеса в сфере информационных технологий,
Национальный исследовательский университет «Высшая школа экономики»;
главный инспектор Счетной палаты Российской Федерации

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: kotelnikovapolina@gmail.com

В статье представлена модель инвестиционного проекта в секторе телекоммуникаций на условиях государственно-частного партнерства (ГЧП), позволяющая связать основные параметры проекта (срок, объем инвестиций, тариф, спрос) с ожидаемыми показателями эффективности. При рассмотрении параметров и показателей эффективности проекта учитываются интересы как государства, так и частной компании.

Алгоритм построения модели и критерии оценки эффективности разработаны на основе стандартов, утвержденных для оценки инвестиционных проектов с долей государственного участия, а именно — Методическими рекомендациями по оценке эффективности инвестиционных проектов, утвержденных Минэкономики РФ, Минфином РФ, Госстроем РФ от 21 июня 1999 г. № ВК 477.

В работе допускается, что для частной компании самым важным критерием оценки эффективности проекта является максимум показателя чистой приведенной стоимости (чистого приведенного дохода) проекта, который характеризуется превышением суммарных денежных поступлений над суммарными затратами для данного проекта с учетом неравноценности эффектов (затрат и результатов), относящихся к различным моментам времени. Таким образом, чем выше данный показатель, тем больший интерес у частной компании вызывает участие в данном проекте.

Представляется, что с точки зрения государства наиболее важными являются два фактора: наличие общественной значимости проекта и минимум затрат государства на реализацию инвестиционного проекта в условиях ограниченной возможности расходования бюджетных средств. Общественная значимость проекта определяется экспертным путем как влияние результатов реализации проекта хотя бы на один из внутренних или внешних рынков: финансовых, рынков продуктов и услуг, рынка труда и т.д., а также на экологическую и социальную обстановку.

Модель позволяет рассчитать различные сценарии для определения оптимальных параметров проекта, обеспечивающих максимальную эффективность при заданном ограничении на объем бюджетных инвестиций.

Результаты расчетов по предлагаемой модели могут быть использованы для принятия органами власти или государственными финансовыми институтами развития решений об участии в проекте. Использование модели и критерии ее оценки предлагается зафиксировать в правилах предоставления субсидии на реализацию таких проектов.

Литература

1. Федеральный закон от 03.12.2012 № 216-ФЗ «О федеральном бюджете на 2013 год и на плановый период 2014 и 2015 годов» // Российская газета. 2012. № 5956, 07.12.2012. С. 19.
2. Государственная программа Российской Федерации «Информационное общество (2011—2020 годы)» // Собрание законодательства Российской Федерации. 2014. №18, 05.05.2014. С. 2159.
3. Коссов В.В., Лившиц В.Н., Шахназаров А.Г. Методические рекомендации по оценке эффективности инвестиционных проектов. М.: Экономика, 2000. 421 с.
4. Инвестиции в отрасль связи и ввод в действие мощностей (2013 г.) [Электронный ресурс]: http://minsvyaz.ru/ru/pages/statistika-otrasli/#section-81 (дата обращения 25.03.2015).
5. Облакова А.В. Имитационное моделирование инвестиционной деятельности в сфере услуг сотовой связи. Дис. ... канд. экон. наук. М., 2009. 203 с.
6. Облакова А.В. Имитационное моделирование инвестиционной деятельности на рынке услуг сотовой связи // Материалы IV всероссийской научно-практической конференции «Имитационное моделирование. Теория и практика», 21-23 октября 2009 г., г. Санкт-Петербург. СПб, 2009. С. 194—198.
7. Telecommunications / Information & Communication Technology PPPs [Электронный ресурс]: http://ppp.worldbank.org/public-private-partnership/sector/telecom (дата обращения 01.04.2015).

FORMATION OF INDIVIDUAL INFOLOGICAL MODELS OF PERFORMANCE MANAGEMENT SYSTEMS

Dmitry V. ISAEV
Associate Professor, Department of Business Analytics,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: disaev@hse.ru

The paper focuses on the questions of infological modeling of performance management systems (PMS), which represent the means of information support of strategic management and help to eliminate the gaps between strategic and operational management levels. Infological models of performance management systems include such elements as information flows, external information objects, functional blocks, functional modules, analytical functions, as well as methods, information systems and processes of management information processing.

It is preferable to develop individual infological models for particular organizations relying on reference models, by individualizing them and detailed elaboration. Among the reference models, there is a basic (generic) infological model that represents the most common characteristics of all the enterprises and organizations, regardless their types and industry affiliation.

The procedure of transition from the basic infological model to an individual model includes four stages.

In the first stage, detailed elaboration of enlarged information flows is performed: each of the enlarged information flows is subdivided into more detailed flows, taking into consideration the peculiarities of a concrete organization. The detailing is provided taking into account types and sources of information, as well its affiliation with particular divisions, business segments and geographical segments. In the second stage, relationships between inputs and outputs of functional modules are discovered. Relying on such relations, preliminary (necessitating additional specification) analytical functions are established. In the third stage, the processes of collection, storage and processing of management information that are available within preliminary analytical functions are defined. Finally, in the fourth stage, the final versions of analytical functions are created by detailing and re-organization of previously defined preliminary functions.

The paper also indicates the possibility of an alternative approach, where developing an individual model starts with the definition of analytical functions.

Introduction

Performance management systems (PMS) represent one of the important classes of management systems. Nowadays they are applied in many large companies and non-for-profit organizations. In the most common sense, such systems may be defined as «formal and informal mechanisms, processes, systems, and networks used by organizations for converting the key objectives and goals elicited by management, for assisting the strategic process and ongoing management through analysis, planning, measurement, control, rewarding, and broadly managing performance, and for supporting and facilitating organizational learning and change» [1]. However, performance management systems often are treated in some narrower sense — as means of information support of strategic management which help to eliminate gaps between strategic and operational management levels [2—4].

One of the stages of performance management systems development process is the stage of design, which includes

formation of infological models. The importance of such models is explained by the fact that they are used as a foundation for subsequent detailed design of particular subsystems and for planning their development. In this connection, an approach to the formation of individual infological models seems topical. Any individual infological model refers to a specific entity and should take into consideration its characteristics — both sectoral and individual.

1. Reference and individual infological models of a performance management system

The methodology of infological modeling of performance management systems has many similarities with other modeling approaches, but at the same time is oriented to a particular task. Such methodology reflects all the material aspects that are important for subsequent design of subsystems. On the other hand, the methodology allows us to avoid excessive detailing, which may reduce understandability of the models for wide range of stakeholders, and first of all, — for managers and specialists who work in the relevant management areas.

The basic terms of the performance management systems modeling methodology include [5]:

✦ information flow — a distinguishable part of management information that is to be transmitted between, at least, two information objects — the source of the information and its recipient;

✦ external information object — a source or a recipient of information, which is situated outside the performance management system;

✦ functional block — an enlarged information object of the performance management system;

✦ functional module — an information object of the system, which is situated within a functional block;

✦ analytical function — an information object of the performance management system which is situated within a functional module. Each analytical function has its owner — a department that supervises its execution;

✦ method of information processing — a way of transforming an input of an analytical function into its output;

✦ information system — hardware and software that provide practical application of an appropriate information processing method;

✦ process — a set of operations that are necessary to transform an input of an analytical function into its output. Each process has its owner, as well as performers of certain operations.

The formation of individual (i.e. related with concrete organizations) infological models of performance management systems may be executed either «from scratch», or relying on typical (reference) models reflecting characteristics of a certain range of management objects.

Within possible reference models, a basic (generic) infological model of the performance management system may be considered to be the most important. The model represents common characteristics of all the organizations, regardless of their types and industry affiliation. The basic model uses a restricted range of modeling elements — information flows, external information objects, functional blocks and functional modules. In particular, the model includes five enlarged external information objects, four functional blocks, thirteen functional modules, as well as information flows between them [6].

The basic model has an enlarged nature: it covers neither detailed analytical functions, nor data processing methods, systems and processes. Furthermore, information flows and external information objects are also enlarged and are subjects of subsequent detailed elaboration in the individual infological models.

Besides the basic infological model, industry-focused reference models may be applied. The industry-focused models reflect specific characteristics of certain industries or types of organizations. Such models are more detailed in comparison with the basic model, because they deal with detailed information flows and external objects.

As to individual infological models, they should contain all the elements of modeling methodology, including detailed information flows, detailed elements of external information objects, as well as analytical functions (with their owners), methods, information systems and processes (with their owners and executors of their operations).

2. Transition from the basic infological model to an individual model

The formation of an individual infological model of a performance management system relying on the basic (generic) model is related to the execution of four stages.

Stage 1. Detailing of enlarged information flows. At this stage, each of the enlarged information flows of a basic model is represented as a set of more detailed flows that represent specific features of concrete organizations. Under the detailing such aspects as the type of information, as well as its affiliation with certain divisions, business segments and geographical segments should be taken into consideration.

The type of information characterizes its nature and its role in the management system. In many cases, types of

information may be determined relying on the applied methods and models of economic analysis and management, for example, environmental analysis models, balanced scorecard methodology, decision making methods, budgeting theory, financial consolidation standards, etc. Often it is convenient to represent the classification of information types as a hierarchy, using «from the general to the particular» logic. Taking into consideration the objective differences in the nature and purpose of the information, different types of information should be represented by different detailed information flows (although one information flow may be used by several recipients).

By a division with which one or another information flow is affiliated, we imply a large element of an organizational structure — a strategic business unit, an enterprise (within a group of companies), or a large department. If the information refers to different divisions, it should be separated into different detailed information flows. The exception is the case when information processing is centralized and the results are delivered simultaneously to several divisions.

A business segment is a component of an entity that provides a particular product, or a group of related products. Identification of business segments may rely on such criteria as the nature of a manufactured product, features of production process, types of customers and methods of delivery, as well as the specificity of the regulatory environment.

Affiliation with a business segment may have an influence on applied types of information, because different businesses need management information of different nature. For example, technological information is important for all the business segments; however, different business segments use information about different technologies. This means that some types of information are relevant to certain business segments and at the same time irrelevant to the others. If information types are represented as a common hierarchical classification, discrepancies may arise on any of the levels. It is also possible to use an approach where particular classifiers are developed for each of business segments separately. At the same time, some types of information are equally relevant to all the business segments (for example, general legislation).

Information relating to different business segments should be represented by different detailed information flows; it is provided by relevancy of information types to certain business segments.

A geographical segment is a component of an entity that provides products or services within a particular economic environment which is inherent to a particular re-

gion. The distinction of geographical segments may rely on such considerations as features of economic and political conditions in regions, relevant risks, as well as relationships between operations in different regions.

Affiliation with geographical segments is also topical for some types of information. For example, marketing information is important for all the geographical segments, however, in different segments information about different regional markets will be used. Affiliation with a geographical segment does not affect the applied types of information: similar operations in different geographical segments require information of the same types. As in the case of business segments, some types of information are equally relevant to all the geographical segments (for example, technological information).

Information relating to different geographical segments should be represented by different detailed information flows.

Stage 2. Establishing linkages between inputs and outputs of functional modules, determining preliminary analytical functions. As for outgoing information of functional modules is created relying on incoming information, at this stage correspondence between inputs of modules and their outputs is established. For each output, a search of inputs whose information (taking into account its processing) is necessary for this output is arranged. Moreover, each output should correspond to at least one input, and each input should be linked with at least one output.

Particular processes of incoming information processing are not considered at this stage. However, methods and appropriate information systems, which will be used to transform inputs into outputs, are to be defined. Determination of methods and data processing systems is extremely important at this stage, because they in many respects have an effect on correspondence between inputs and outputs.

When establishing correspondence between inputs and outputs of functional modules, two types of problem situations may arise.

In the first case, for some of the outputs, determining appropriate sets of inputs (so that information of these inputs will be sufficient for formation of the outputs) appears unsuccessful. The cause is insufficiency of the incoming information. In this case, it is necessary to revise the scope and content of the incoming information flows, and then — to introduce some new information flows or to amend content of existing ones.

In the second situation, for some of the outputs aggregated incoming information arriving from corresponding inputs is excessive, but still all these inputs are really necessary and none may be excluded from considera-

tion. Such a situation may arise from a lack of information flow detailing. In this case, it is necessary to arrange additional detailed elaboration of incoming information flows which are excessive regarding certain outputs.

After establishing the correspondence between inputs and outputs of functional modules, relying on these linkages, analytical functions are formed. Each of these functions, as usual, is related to one particular output of the functional module: each function receives incoming information that, after certain processing, is transformed into outgoing information for the particular output of the functional module.

The exceptions are cases when several outputs are created as a result of joint information processing. In such cases, an analytical function may be related simultaneously with several outputs.

The scope of analytical function determined at this stage is not final, because they were established just relying on linkages between inputs and outputs, regardless of concrete processes of information collection, storage and processing. Furthermore, one has to define more precisely the scope of the function and their content, taking into account particular processes and their owners. Therefore, the analytical functions determined at this stage may be treated as preliminary functions.

Stage 3. Determining processes of management information gathering, storage and processing within preliminary analytical functions. At this stage, within preliminary analytical functions, we describe processes of transformation of these functions' inputs into their outputs. When describing the processes, previously defined data processing methods and information systems are taken into consideration. The processes are described using existing notations such as ARIS, IDEF 3 or BPMN. In a parallel way, the methods and information systems are elaborated in more detail. For all the defined processes (or their considerable fragments), their owners (departments and managers responsible for execution of the processes) are appointed.

At this stage, cases of duplication of the processes (i.e. execution of the same processes within different analytical functions), as well as situations when processes of one analytical function have different owners are quite probable. Such situations arise because of the peculiarity of the procedure of determining preliminary analytical functions and they should be adjusted in the next, final stage.

Stage 4. Determining final analytical functions. The fourth and final stage is associated with redesign of preliminary analytical functions and determining final versions of the functions. Moreover, the preliminary func-

tions are analyzed from the point of view of duplication of processes. All the processes and their considerable parts, which are duplicated within one or several preliminary functions, are to be excluded from these functions. Instead, for each such process (fragments) a new analytical function is created. As a result, preliminary analytical functions are transformed into new functions that are free from duplication of processes.

After elimination of the process duplication effect, analytical functions are analyzed from the point of view of the existence of multiple owners of the processes which are available within them. Such situations are undesirable because the existence of multiple owners within the same function may appear critical for appointment of the owner of the analytical function. Therefore, in such cases it is recommended to split the analytical function into a few new functions, so that all the processes within any of the new functions have only one owner (that becomes the owner of the whole function).

The combination of different processes owners within one analytical function is acceptable as an exception if the vast majority of the processes within the function are associated with the same owner. In this case, such an owner is appointed as the owner of the whole analytical function, while «minority» owners act as its subcontractors.

At the fourth stage, the direct performers of all the operations of the processes are also appointed. Generally, the performers of certain operations may differ from the owners of processes and analytical functions.

3. An alternative approach

As an alternative to the procedure of transition from the basic infological model to an individual model described above, an approach starting with analytical function determination may be considered. The scope and content of analytical functions should be determined relying on two criteria — commonality of operations performed within one or another function, and the possibility to appoint a certain owner who is able to take responsibility for execution of the function. For each determined analytical function, its inputs and outputs, as well as the methods of management information processing and appropriate information systems should be defined.

Further steps are focused on determining information linkages between analytical functions and, for external information flows — between analytical functions and external information objects. This means that each input of an analytical function should be linked with a single source — an output of another function or external object, and each output — with outputs of other functions

or external objects. For this purpose, first of all, correspondence of inputs and outputs of analytical functions with functional modules and external objects (information sources or recipients of the function) is performed. At the same time, enlarged information flows related to inputs and outputs of analytical functions are defined. These steps allow us to link analytical functions and, if necessary, external objects and, consequently, to determine detailed information flows.

Using this approach, some problems related with identification of linkages between inputs and outputs of analytical functions may arise. It is also possible that some of the information requirements of external objects – recipients of information from the performance management system appear unfulfilled. The cause of such situations may be related to lack of content of information flows, or even to missing some information flows that really are necessary. Such discrepancies may arise when determining inputs and outputs of analytical functions. Adjustment of problem situations arising during the alternative approach application may require either revision of content of the information flows (and, consequently, inputs and outputs of analytical functions), or addition into the model of some new information flows or new analytical functions.

Adjustment of the model may have an iterative nature and may require considerable investment of the time of managers and specialists. Nevertheless, the advantage of the alternative approach is the possibility of better use of knowledge, experience, initiative and enthusiasm of employees who work on lower levels of the management hierarchy, while determining the ways of performing managerial functions more efficiently. In contrast to the main approach described above («from the requirements»), the alternative approach may be considered as an approach «from possibilities». It may be concluded that both approaches may be applied in the practice of modeling performance management systems.

Conclusion

Infological modeling is extremely important for designing performance management systems and, in the general context, – for managing the development of such systems. The infological model is based on such elements as information flows, functional blocks, functional modules, analytical functions, as well as methods, information systems and processes of management information processing.

While the basic (generic) reference infological model of a performance management system represents the most common characteristics of organizations of all types and industries, individual models are designed for particular organizations. The most efficient way of forming individual infological models is transition from the basic reference model by its detailed elaboration. The procedure of an individual model formation relying on the basic model comprises several stages, including detailing of enlarged information flows, establishing linkages between inputs and outputs of functional modules with subsequent determination of preliminary analytical functions, determining processes of management information gathering, storage and processing within preliminary analytical functions, and determining final analytical functions.

There is also an alternative approach. According to this approach, the first step of the individual infological model formation process is definition of analytical functions. Then correspondence between analytical functions and external information objects are established and detailed information flows are determined. The alternative approach allows us to use competences of management personnel more efficiently and, in spite of some objective disadvantages, may also be applied in modeling practices of performance management systems. Detailed elaboration of stages of the alternative approach, as well as related problem situations, may be considered as a subject of further research. ∎

References

1. Ferreira A., Otley D. (2009) The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, no. 20, pp. 263–282.
2. Cokins G. (2004) *Performance management: Finding the missing pieces (to close the intelligence gap)*. Hoboken, NJ: John Wiley & Sons.
3. Coveney M., Ganster D., Hartlen B., King D. (2003) *The strategy gap: Leveraging technology to execute winning strategies*. Hoboken, NJ: John Wiley & Sons.
4. Eckerson W.W. (2006) *Performance dashboards: Measuring, monitoring and managing your business*. Hoboken, NJ: John Wiley & Sons.
5. Isaev D.V. (2013) Proektirovanie sistem informatsionnoi podderzhki korporativnogo upravleniya i strategicheskogo menedzhmenta [Design of corporate governance and strategic management information support systems]. *Audit and Financial Analysis*, no. 6, pp. 329–334.
6. Isaev D. (2012) Performance management information support system: A conceptual model. *European Journal of Economics, Finance and Administrative Sciences*, no. 52, pp. 6–20.

ФОРМИРОВАНИЕ ЧАСТНЫХ ИНФОРМАЦИОННО-ЛОГИЧЕСКИХ МОДЕЛЕЙ СИСТЕМ УПРАВЛЕНИЯ ЭФФЕКТИВНОСТЬЮ

Д.В. ИСАЕВ
кандидат экономических наук, доцент кафедры бизнес-аналитики,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: disaev@hse.ru

Статья посвящена вопросам информационно-логического моделирования систем управления эффективностью (performance management systems, PMSs), представляющих собой средства информационной поддержки стратегического менеджмента, помогающие устранить разрывы между стратегическим и текущим уровнями управления. Информационно-логические модели систем управления эффективностью включают такие элементы, как информационные потоки, внешние информационные объекты, функциональные блоки, функциональные модули, аналитические функции, а также методы, информационные системы и процессы обработки управленческой информации.

Формирование частных информационно-логических моделей для конкретных организаций целесообразно осуществлять на основе референтных моделей, путем их конкретизации и детализации. К числу референтных относится базовая (обобщенная) информационно-логическая модель, отражающая наиболее общие характеристики всех предприятий и организаций, независимо от их типов и отраслевой принадлежности.

Процедура перехода от базовой информационно-логической модели к частной предусматривает четыре этапа.

На первом этапе осуществляется детализация укрупненных информационных потоков: каждый из укрупненных информационных потоков базовой модели разбивается на более детальные потоки, учитывающие особенности конкретной организации. Детализация осуществляется с учетом типов и источников информации, а также ее принадлежности к определенным подразделениям, бизнес-сегментам и географическим сегментам. На втором этапе выявляются связи между входами и выходами функциональных модулей, на этой основе формируются предварительные (т.е. нуждающиеся в дополнительном уточнении) аналитические функции. На третьем этапе производится определение процессов сбора, хранения и обработки управленческой информации, которые протекают в рамках предварительных аналитических функций. Наконец, на четвертом этапе происходит формирование окончательных аналитических функций, путем детализации и реорганизации определенных ранее предварительных функций.

В работе также указывается на возможность альтернативного подхода, при котором построение частной модели начинается с определения аналитических функций.

Литература

1. Ferreira A., Otley D. The design and use of performance management systems: An extended framework for analysis // Management Accounting Research. 2009. No.20. P. 263-282.
2. Cokins G. Performance management: Finding the missing pieces (to close the intelligence gap). Hoboken, NJ: John Wiley & Sons, 2004. 304 p.
3. The strategy gap: Leveraging technology to execute winning strategies / M. Coveney [et al.]. Hoboken, NJ: John Wiley & Sons, 2003. 240 p.
4. Eckerson W.W. Performance dashboards: Measuring, monitoring and managing your business. Hoboken, NJ: John Wiley & Sons, 2006. 318 p.
5. Исаев Д.В. Проектирование систем информационной поддержки корпоративного управления и стратегического менеджмента // Аудит и финансовый анализ. 2013. №6. С. 329—334.
6. Isaev D. Performance management information support system: A conceptual model // European Journal of Economics, Finance and Administrative Sciences. 2012. No. 52. P. 6—20.

RESOURCE CHARACTERISTICS OF WAYS TO ORGANIZE A DECISION TREE IN THE BRANCH-AND-BOUND METHOD FOR THE TRAVELING SALESMEN PROBLEM

Mikhail V. ULYANOV
Leading Researcher, V.A. Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences;
Professor, Software Management Department,
National Research University Higher School of Economics
Address: 65, Profsouznaya Street, Moscow, 117997, Russian Federation
E-mail: muljanov@mail.ru

Mikhail I. FOMICHEV
Student, Software Engineering Program,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: mikhail.fomichev94@gmail.com

The resource efficiency of different implementations of the branch-and-bound method for the classical traveling salesman problem depends, inter alia, on ways to organize a search decision tree generated by this method. The classic «time-memory» dilemma is realized herein either by an option of storing reduced matrices at the points of the decision tree, which leads to reduction in the complexity with additional capacity cost, or matrix recalculation for the current node, which leads to an increase in complexity while saving memory. The subject of this paper is an experimental study of temporal characteristics of solving the traveling salesman problem by the branch-and-bound method to identify a real reduction of span time using additional memory in a selected structure of a decision tree. The ultimate objective of the research is to formulate recommendations for implementing the method in practical problems encountered in logistics and business informatics.

On the basis of experimental data, this paper shows that both considered options of the classic algorithm for the traveling salesman problem by the branch-and-bound method generate software implementations with an exponential dependence on the execution time of the input length. The experimental results permit us to suggest that the applicability of an additional memory capacity of no more than 1 GB results in a significant (up to five times) reduction of the time span. The estimate of the resulting trend makes it possible to recommend practical application of the software implementation of the branch-and-bound algorithm with storage of matrices - with a really available 16 GB random-access memory and with limitation of the expected average computation time of about one minute on modern personal computers whereby problems having a dimension no more than 70 can be solved exactly.

Introduction

A large variety of practical settings in the area of business informatics and logistics comes down to the classical traveling salesman problem. The abundance of heuristic methods for its solution does not mean rejection of obtaining exact solutions for this problem. Obviously, for precise methods having an exponential complexity the researchers would like to have estimates of the dimension of problems which can be solved within a reasonable time, as well as modified exact algorithms with a better time efficiency.

For certain algorithms, an increase of time efficiency can be achieved by using an additional amount of available memory. This situation arises in algorithms allowing us to replace re-computations for storing previously obtained intermediate results. This approach can also be used for algorithms implementing the branch-and-bound method for the traveling salesman problem. By the example of modifications of the classical algorithm for the exact solution of the traveling salesman problem using the branch-and-bound method proposed by J.D.C.Little, K.G.Murty, D.W.Sweeney and C.Karel [1], this paper illustrates a possible increase of the time efficiency using an additional memory to store matrices in the data structure of the search decision tree.

1. Setting up the traveling salesman problem

The problem description lies in the fact that the salesman has to visit a number of cities for product advertising and selling purposes. It is assumed that between each pair of cities there is an internal transportation. Let us call these visiting rules a tour in which each city is visited only once. The fare between the cities is known, and, in general, the return fare is different. The challenge is to find the traveling salesman tour with minimal cost. It is obvious that the number of tours is finite, and we can solve the problem by straightforward enumeration. But, unfortunately, a complete set will contain $(n-1)!$ tours, for this reason an exhaustive solution algorithm for solving the real dimensions of the problem becomes totally unacceptable.

Setting up the problem in terms of the graph theory is built by association of cities with the graph nodes, and the transport routes and fare with loaded arcs. We get a complete directed asymmetric graph at n nodes without proper loops prescribed by the cost matrix $C = (c_{ij})$. The problem is called symmetric if the fare between any two cities does not depend on the direction, and asymmetri-

cal, if that is not the case. The absence of proper loops can be denoted as $c_{ii} = \infty \ \forall i = \overline{1,n}$. Then, the problem is formulated as a problem of finding a covering (complete) cycle at the lowest cost, which is called a tour, on a complete directed graph specified by asymmetric (in the general case) cost matrix C [2].

Let us also enter the problem definition in Euclidean integral-valued space E_z^{n-1} [3]. It is based on the concept of permutation on a set of integers. Let us further represent through $\pi(k,l), l \geq k$ a set of all permutations of integers (it is obvious that a number of such permutations is $(l-k+1)!$) and consider a set of permutations $\pi(2,n)$, which contains $|\pi(2,n)| = (n-2+1)! = (n-1)!$ of various permutations (exactly the number of various tours is available in the traveling salesman problem with n cities). Since a tour is a complete cycle across the nodes, the initial node of the tour can be chosen arbitrarily. Let us fix a node number 1, and assume that some permutation $\mathbf{x}$ from a set $\pi(2,n)$ orders the tree traversal starting from the first, and after the last point given by this permutation, we return to the beginning of the tour. Therefore, the components of vector $\mathbf{x}$ in space E_z^{n-1} are simply ordered numbers from two to n, and the vector is associated with some permutation from $\pi(2,n)$, and it can be written:

$$\mathbf{x} \in \pi(2,n), x_i \in \{2,...,n\}, i = \overline{1,n-1}, x_{ij} \neq x, \text{ with } i \neq j \quad (1)$$

Let us note that the sum of squares of the components of different vectors $\mathbf{x}$ is the same: these are different permutations of numbers from two to n. Thus, the end points of various vectors $\mathbf{x}$ are points on the positive hemisphere in E_z^{n-1} centered at the origin and with a radius of

$$r = \sqrt{\sum_{i=2}^{n} i^2} = \sqrt{\frac{n \cdot (n+1) \cdot (2n+1)}{6} - 1}, \ \mathbf{x} \in S_z^{n-1}(0,r).$$

Let us further determine the function

$$N \times N \underset{c}{\to} R^+, (i,j) \to c(i,j), (i,i) \to \infty$$

which assigns the value of the edge to each ordered pair of graph nodes, incident to this pair. In the constructed formalism, the traveling salesman problem in space E_z^{n-1} has the following statement:

$$\begin{cases} f(\mathbf{x}) = c(1,x_1) + \sum_{i=1}^{n-2} c(x_i, x_{i+1}) + c(x_{n-1}, 1), f(\mathbf{x}) \to min \\ \mathbf{x} = (x_1, \cdots, x_{n-1}) \in \pi(2,n) \end{cases} \quad (2)$$

The branch-and-bound method under further consideration works, if not explicitly, with this statement of the traveling salesman problem.

2. Description of branch-and-bound method for the traveling salesman problem

The general idea of the branch-and-bound method is a separation of the entire set of feasible solutions into subsets to further reduce the enumeration - a branching procedure. Every such subset shall be linked to an estimate (lower bound in the minimum search), providing a truncation of those subsets which intentionally do not contain an optimal solution − this is a procedure for constructing bounds. Therefore, the method results in investigation of a tree solution space model. A set of all salesman tours in which the objective functional (2) is minimized, specifying the tour cost, is such an initial subset in the problem under investigation. The ideas presented below from the algorithm authors [1] are a kind of classic of the branch-and-bound method. To construct the algorithm, two basic procedures, i.e. branching and bounding, are proposed. Let us consider the branching process [1]. The construction of a search decision tree starts with the root, which will correspond to a set of all tours, i.e., a root of the tree is set R of all $(n - 1)!$ tours possible tours in the problem with n towns. The branches going out from the root are determined by selecting one edge, for example, arc (k, l). The idea of the algorithm authors [1] is to divide the current set of tours into two subsets: one which most likely contains the optimal tour, and the other one which most likely does not contain this tour. To do this, a special algorithm for selecting arc (k, l) is proposed, which likely is a part of the optimal tour. The set R is divided into two subsets $\{k, l\}$ and $\{\overline{k,l}\}$. The subset $\{k, l\}$ includes all tours from R containing arc (k, l), i.e. passing through it, and the subset $\{\overline{k,l}\}$ includes tours that do not contain this arc. Let us note that the idea of the algorithm authors is very promising if the branching process is organized so that at each step the «right» edge is selected, the entire process will be completed after n steps. A fragment of such tree is shown in *Fig. 1* (the figure is taken from paper [2]).
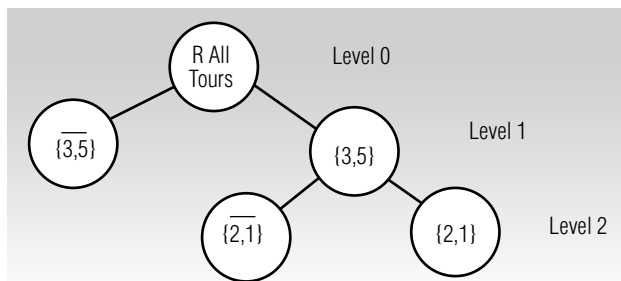
Fig. 1. Fragment of a search decision tree

The lowest cost limit of any tour of this set is associated with each tour of this set. It is obvious that the problem is to obtain the most accurate lower bounds.

The reason for that is as follows. Let us assume that a specific complete tour T with cost $s(T)$ has already been obtained. If the lower bound associated with a set of tours represented by a search tree node is higher than $s(T)$, then until the end of the search process this and all subsequent points are not to be considered. In implementation this leads to truncation of the search decision tree by shedding all leaves of the search tree having a value higher than $s(T)$. A detailed presentation of other steps of the method can be found, for example, in papers [2, 3].

3. Diagram of the branch and bound method for the traveling salesman problem

Let us present the following diagram of the branch-and-bound method (B&B) for the traveling salesman problem, in which the following notations are introduced [2]. Let X be the current top of the search tree, and (k, l) be an edge across which the branching occurs. Let us denote the tops immediately following X through Y and $\overline{Y}$. The set Y is a subset of tours from X passing via arc (k, l), and set $\overline{Y}$ is a subset of tours from X not passing via arc (k, l). Let us denote the calculated the lowest bounds for sets Y and $\overline{Y}$ by $w(Y)$ and $w(\overline{Y})$ respectively. Let us denote the cheapest tour known to the algorithm at the given moment by z_o, provided that at the time of initialization is $z_o = \infty$ [1].

A(C, n) Diagram of the branch and bound method for the traveling salesman problem
(n - dimensions, C – cost matrix.)
1. Initializaiton.
2. Reduction of cost matrix C.
3. Setting a root of the search decision tree X=R
Reduction of the initial matrix – calculation w(X).
While (w(X) < z0)
    begin
        4. Selection of branching node (k, l).
        5. Branching process. Creating top Y⁻
        and calculation w(Y⁻).
        6. Branching process. Creating top Y
        and calculation w(Y).
        If (the order of cost matrix in top Y = 2)
            then
                begin
                    7. Performance of exhaustive estimation for top Y
                    If (w(Y) < z0)
                        then
                            begin
                              z0 = w(Y)(store the tour)
                        end
            end
        8. Selection of the following point of the search decision tree and setting X
        9a. Calculation of matrix fragment C, corresponding to selected top X, based on a route from the root to the current point.
(or)
        9b. Reading of matrix fragment C, corresponding to selected top X, from the structure of storage of the search solution tree.
    end (while w(X) < z0)
Optimal solution with cost z0 is not found.
End.

For the purposes of this article, let us draw our attention to some stages in this enlarged scheme.

Step 1. Selection of a data structure for storing the search decision tree is of interest here. At the same time, it is necessary to bear in mind that a number of the search tree points can be significant.

Step 3: Initialization of the search tree root. The main question is whether the cost matrix will be stored together with the top of the tree. An alternative option is to re-calculate the cost matrix for the current top based on the original one. This is a classical choice between performance and memory requirements. The classical algorithm requires recalculation of the matrix costs for the newly determined current point of the search decision tree.

Step 9. Since the node of the search tree *X* in step 8 has already been selected as a current one, the task of this step is to obtain a cost matrix corresponding to top *X*. If we keep the truncated cost matrices together with the nodes of the search tree, the matrix already exists (9b). Otherwise, we need to find a path from the root to this node and consistently adjust the original cost matrix (9a), i.e. re-compute it for this point.

Discussion of algorithm specifics. What results in terms of complexity can be expected for different cost fixed-dimension matrices? It is known and suggested that the complexity of all more or less efficient algorithms implementing the branch-and-bound method for the traveling salesman problem in the worst case is exponential. This assumption is based on the fact that the traveling salesman problem is *NP*-hard and, therefore, any exact algorithm of its solution has an above-polynomial complexity. In the best case, if the dimension of the cost matrix keeps declining, the estimate of complexity is polynomial. This is obvious, because the assessment of complexity of each internal step (4-9) of the polynomial algorithm by a linear dimension *n* of the cost matrix, and, in the best case, the main loop is executed no more than *n* times, because the tour consists of *n* arcs. Therefore, the spread of expected execution time with a fixed dimension of the cost matrix *n* is very high and depends on the numerical values of its elements. This is an example of quantitative parametric algorithm, with a strong parametric dependence; the algorithm belongs to the class *NPRH* [3]. Theoretical analysis of the expected complexity for a particular entry based on a preliminary study of the cost matrix is very complicated and often goes beyond the scope of our analytical capabilities.

However, the time efficiency will be determined specifically as a selected structure to store the search de-

cision tree, and a decision in step 9 with storage or re-computation of the current cost matrices. The matter is that an average number of active nodes grows exponentially. The experimental data from [4] gives an experimentally obtained approximation $R(n) = 3,6932 \cdot e^{0,1819n}$. In this regard, the effectiveness of elementary operations with the decision tree (add, delete, select) determines the time effectiveness of the software implementation of the algorithm as a whole [5, p. 130-131].

4. Data structures for the search decision tree

In order to select the data structures for the search decision tree, let us look at the three best-known options of the tree organization [6, 7], i.e. a binary heap, red-black tree and AVL tree

Table 1 shows the complexity of operations (information is taken from [6, 7]), which are performed above the search decision tree in the branch-and-bound method for the traveling salesman problem. It is evident from the table that the operating time (asymptotically) of the insert and element removal operations for all three data structures are equal. However, the minimal element search operation for the binary heap is performed over a constant time, but for the AVL and red-black trees over a logarithmic time. Moreover, in the functions of complexity of these operations coefficients with the term with higher exponent of the polynomial are much higher for red-black and, especially, AVL trees than for the binary heap, because additional operations are spent on the tree balance. It is also worth noting that three data structures have the same requirements for storage $O(n)$. Therefore, it may be assumed that to store decision tree leaves a binary bunch will be suitable structure as compared to others, although this issue will be studied additionally and is beyond the scope of this study.

Table 1.

Complexity of operations for various ways of tree organization

Reference line	Binary heap	Red-black tree	AVL-tree
Complexity of insert	$O(log(n))$	$O(log(n))$	$O(log(n))$
Complexity of remove	$O(log(n))$	$O(log(n))$	$O(log(n))$
Complexity of search minimal element	$O(1)$	$O(log(n))$	$O(log(n))$

On the basis of the data presented by the authors, a decision was taken in this study to use a binary heap as the structure for storing the search decision tree.

5. Objectives
of the experimental research

In this experimental research, the authors set the following objectives. The main objective is to illustrate a possible increase of the time efficiency through the use of additional memory for storage matrices in the search decision tree structure. The issue concerning the effect of matrix storage in the nodes of the search tree on the exponent in a trend of the average calculation time or only on a multiplicative constant in the trend function is of additional interest. Secondary objectives were to study a dependence of the memory capacity on the entry length in the comparative analysis of experimental results and the data obtained in paper [4], the distribution of the relative frequencies of the observed times, and to obtain data on the scope of variation and sample standard deviations for characteristics under study and design average time prediction for longer entry lengths.

6. Description of the experiment
plan and hardware

Consideration was given to an asymmetric traveling salesman problem. The cost matrix was generated by a pseudorandom uniform generator standard for C++ pseudo-uniform generator. In order to reduce the total time of the experimental research, the cost matrix elements had an integer type, and a range of generation of oriented edge weights was chosen from 1 to 2^{15}. For the same purposes, the range of values of input lengths of the traveling salesman problem (linear dimension of the matrix values) from 25 to 45 with an increment of 1 was selected.

Algorithm [1] was implemented (with no prediction of the final tour of the greedy algorithm) with the structure of storing the search decision tree in the form of a binary heap in the classical implementation with recalculation of the cost matrix (i.e., without using an additional memory) and in a modified implementation with storage of a local cost matrix (for nodes of the tree).
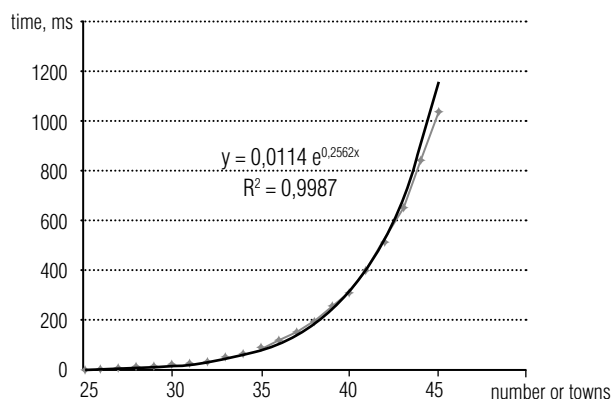
For each fixed input length, 10,000 pseudo-random cost matrix generations were performed, for each of which a classic algorithm implementation was started (with recalculation of matrices) and modified algorithm implementation (with matrix storage) [1]. The run-time of the algorithm software implementation, generated number of nodes of search decision tree was measured for each start, and a maximum value of additional memory spent for matrix storage was also measured for modification with matrix storage.

Experiments were carried out on a stationary machine with the following characteristics:

✦ Processor: Intel i7 3770K 3800 GHz;

✦ Random access memory: Kingston KHX-1600C9D3P1 16 GB;

✦ Motherboard: GIGABYTE GA-Z77X-D3H;

✦ Operating system: Fedora 21 workstation.

The algorithms are implemented in language C++. The compiler version: gcc 4.9.2 20150212 (Red Hat 4.9.2-6) (GCC). Let us note that in this regard the total time of computational experiment (10,000 starts for 21 values of the input length) was about 20 hours.

7. Results and discussion

Within the main objective of the study the authors obtained the results shown in *Fig. 2* and *3*. For both options of organization of the decision tree (matrix storage or recomputation) the observed average time grows exponentially. At the same time, we have to note that the option

Fig. 2. Dependence of an average time of tour calculation
without additional memory on the input length
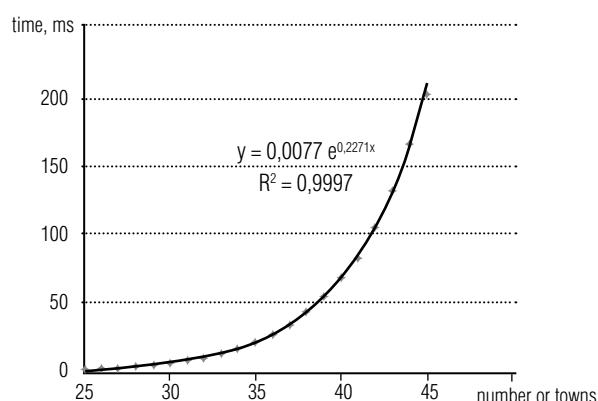
Fig. 3. Dependence of an average time of tour calculation
on the input length with the additional memory
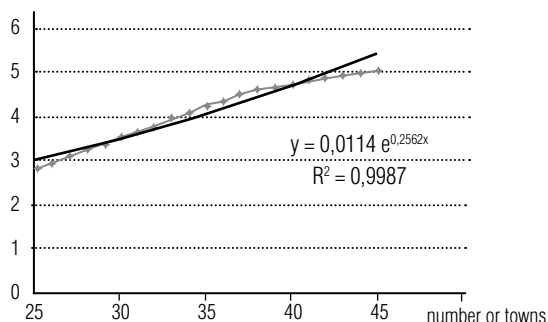
Fig. 4. Dependence of a relation of the average time of tour calculation without additional memory to the average time of tour calculation with the additional memory on input length

with matrix storage results in a change of not only the multiplicative constant, but also the exponent.

From the trend of experimental results (*Fig. 2* and *3*) it follows that with an increase of the input length of the problem the observed reduction of time $d(n)$ (time ratio) will grow in trend $d(n) = 1,478 \cdot e^{0,0291n}$. The results and trend $d(n)$ are provided in *Fig. 4*.

It is evident that the effect of matrix storage will be higher, the deeper the search tree, as a «long» path from the root leads to more complicated recalculations for the original matrix. Our experimental data shows that the use of additional memory reduces the exponent from 0.2562 to 0.2271, i.e. by 0.0291. For the problem with 45 cities this reduction was 5.087 times (from 1.04 to 0,204 sec), all of which indicates considerable option efficiency with additional memory. The resulting trend for $d(n)$ suggests that the observed reduction of time will also increase with the increase of the input length. The resulting prediction is given below.

The following results are obtained for additional goals of the research.

The results on dependence of additional memory capacity on the input length are shown in *Fig. 5*. For input length 45 the additional memory requirements average 30.71 MB, providing a 5-fold reduction of time.
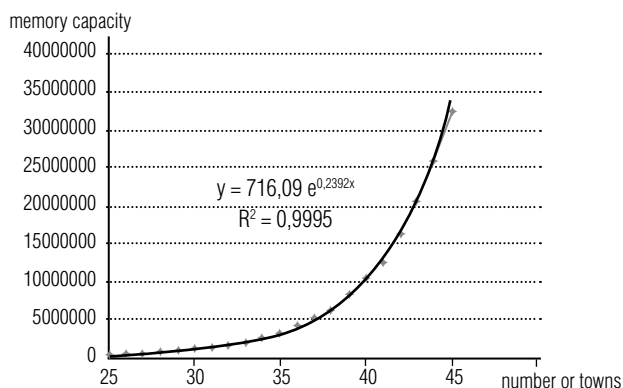
Fig. 5. Dependence of the average capacity of additional memory on the input length

By comparative analysis of the experimental results and data obtained in paper [4] on dependence of the average number of nodes of the search decision tree on the input length, the results obtained by the authors (*Fig. 6*) give the following trend of estimate of an average number of nodes of the decision tree: $R(n) = 5,323 \cdot e^{0,1831n}$ that is qualitatively comparable with the results from [4] — $R(n) = 3,6932 \cdot e^{0,1819n}$. By the exponent, a disagreement with our data is not more than 0.65%.
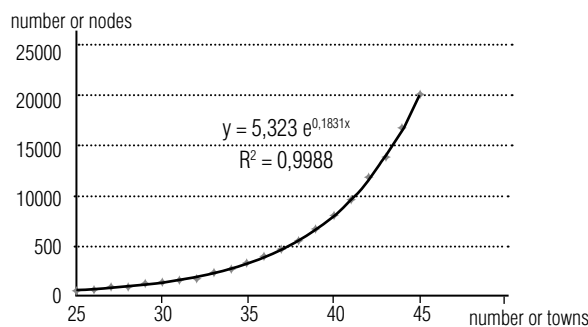
Fig. 6. Dependence of the average number of nodes of the search decision tree on the input length

The discrepancy in the multiplicative constant can be attributed to lower values in a range of studies in paper [4] — from 10 to 30 and may be specifics of the initial data generation. The exponential growth of a number of generated tops of the search decision tree defines the exponential nature of time and capacity characteristics of the software implementations of the method.

Based on the distribution of relative frequencies of the observed time, the following interesting results are obtained which are shown in *Fig. 7*. We have a pronounced left asymmetry of distribution of relative frequencies, indicating that the major part of the observed times is close to the best case. Large times are rare enough. For the point of the sample mean — 204.25 ms a relative frequency of 0.7138 is summarized, and the sample quantile 0.95 is in point 728.70 ms, beyond which only 500 (5%) out of 10,000 observed times lie with a maximum value of 8888.74 ms.

According to the ranges of deviation and sampling root-mean-square deviations for characteristics being studied, the fact that algorithms implementing B&B belong to a class *NPRH* [3] is clearly confirmed by the data on the range of variability of the observed times: if n = 45 for option without matrix storage, we have a range from 3.38 ms to 54058.1 ms at a sample mean of 1039.08 ms. For option with matrix storage from 2.08 ms to 8888.74 ms at sample mean of 205.25 ms. The same significant range is observed for the additional memory spent at n=45 — from 185 936 to 882 836 484 bytes (from 181.57 Kb to 841.39 Mb) with an average value of 30.71 MB. All these ranges
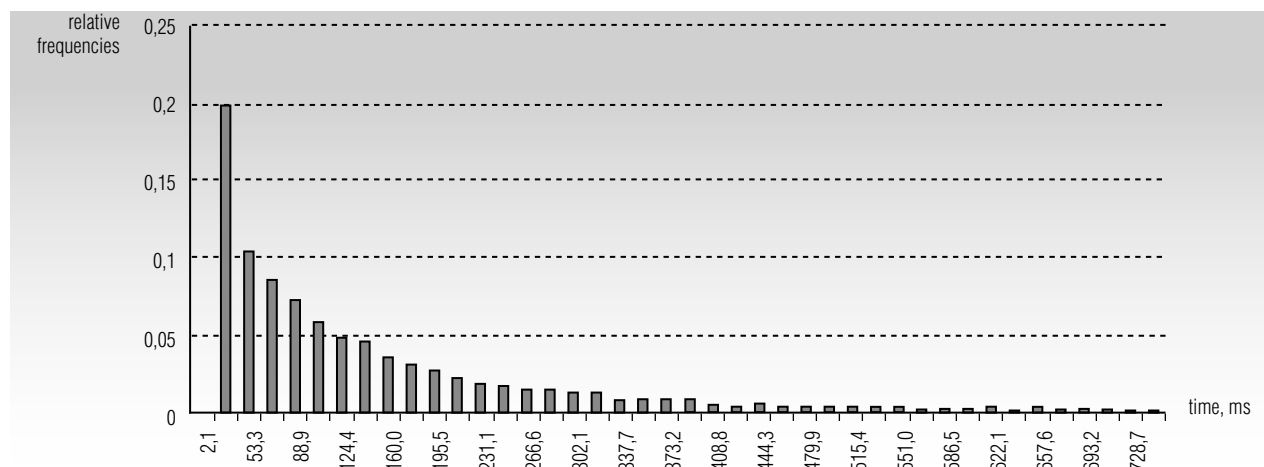
Fig. 7. Relative frequencies of execution times (with matrix storage) for problem
with *n*=45 (first 42 half-segments out of 500 by 10,000 experiment results)

are caused by a range of variability of a number of generated nodes of the search tree − from 111 to 820,964 with an average of 20,121.72. The presence of rare but significantly large runs identified both large sample root-mean-square deviations − 1783.65 ms 343.91 ms for the timed and 49 685 928.5 bytes for the spent memory.

Based on the trend of times and additional RAM spent for storing matrices on tops of the search tree, the authors obtained the following prediction of resource characteristics for problems with a high dimensionality presented in *Table 2*.

Table 2.

Prediction of resource characteristics

N-r of cities	Prediction of tour calculation time without additional memory	Prediction of tour calculation time with additional memory	Prediction of time relation	Prediction of average requirements of additional memory
45	1 sec	0.2 sec	5	30.71 MB
54	7 sec	1 sec	7	172.3 MB
70	11.7 min	1 min	11,7	12.47 GB
80	2.5 hours	10 min	15	136.37 GB
88	19.6 hours	1 hour	19,6	924.26 GB
102	29.5 days	1 day	29,5	25.69 TB

Let us note that, beginning from the input length 72, expected memory requirements of the additional memory already exceed the modern standard volume of RAM by 16 GB.

Conclusion

Thus, based on the experimental research consisting of 210,000 solutions of the asymmetric traveling salesman problem with a random uniform generation of cost matrices for a range of input lengths from 25 to 45, and the processing of the results obtained, the article shows that:

✧ both considered options of the classical algorithm for solving the traveling salesman problem using the branch-and-bound method generate software implementations with an exponential execution time dependence on the input length;

✧ the use of additional memory of an acceptable volume leads to a significant reduction of calculating time (up to 5 times with *n*=45, with maximum memory requirements not exceeding 1 GB, with average requirements of 30.71 MB);

✧ the obtained time distribution indicates a pronounced left asymmetry of skewness of distribution, which leads to «small» execution times expected within probabilistic quantile 0.95 in rare but significant (i.e., by orders of magnitudes exceeding an average value) runs;

✧ software implementation of the algorithm with matrix storage can be practically used for the exact solution of the traveling salesman problem with the input length of no more than 70 with the actually available memory of 16 GB with an expected average computation time of about one minute on modern personal computers. ∎

References

1. Little J.D.C., Murty K.G., Sweeney D.W., Karel C. (1963) An algorithm for the traveling salesman problem. *Operations Research*, no. 11, pp. 972–989.
2. Goodman S.E., Hedetniemi S.T. (1981) *Vvedenie v razrabotku i analiz algoritmov* [Introduction to the design and analysis of algorithms]. Moscow: Mir (in Russian).

3. Ulyanov M.V. (2008) *Resursno-effektivnye komp'juternye algoritmy. Razrabotka i analiz* [Resource-efficiency computer algorithms. Development and analysis]. Moscow: FIZMATLIT (in Russian).

4. Ermoshin A.S., Plisko V.A. (2006) Issledovanie dereva reshenij metoda vetvej i granic v zadache kommivojazhera [Research of a decision tree in branch-and-bounds method for traveling salesman problem]. P*rogrammnoe i informacionnoe obespechenie sistem razlichnogo naznachenija na baze personal'nyh EVM* [Software and information in different systems based on personal computers]. Moscow: MGAPI, no. 9, pp. 76–82 (in Russian).

5. Sygal I.H., Ivanova A.P. (2007) *Vvedenie v prikladnoe diskretnoe programmirovanie: Modeli i vichislitelnie algoritmi* [Introduction to applied discrete programming: models and algorithms]. Moscow: FIZMATLIT (in Russian).

6. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C. (2005) *Algoritmy: postroenie i analiz* [Algorithms: Development and analysis]. Moscow: Williams (in Russian).

7. Virt N. (2010) *Algoritmy i struktury dannyh. Novaja versija dlja Oberona* [Algorithms and data structure. New version for Oberon]. Moscow: DMK Press (in Russian).

РЕСУРСНЫЕ ХАРАКТЕРИСТИКИ СПОСОБОВ ОРГАНИЗАЦИИ ДЕРЕВА РЕШЕНИЙ В МЕТОДЕ ВЕТВЕЙ И ГРАНИЦ ДЛЯ ЗАДАЧИ КОММИВОЯЖЕРА

М.В. УЛЬЯНОВ
доктор технических наук, профессор, ведущий научный сотрудник,
Институт проблем управления им. В.А.Трапезникова РАН;
профессор департамента программной инженерии,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 117997, г. Москва, ул. Профсоюзная, д. 65
E-mail: muljanov@mail.ru

М.И. ФОМИЧЕВ
студент бакалавриата образовательной программы «Программная инженерия»,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: mikhail.fomichev94@gmail.com

Ресурсная эффективность различных реализаций метода ветвей и границ для классической задачи коммивояжера зависит, в том числе, от способов организации поискового дерева решений, порождаемого этим методом. Классическая дилемма «время-память» реализуется здесь либо вариантом хранения усеченных матриц в вершинах дерева решений, что приводит к сокращению трудоемкости при дополнительных емкостных затратах, либо перевычислением матрицы для текущей вершины, что ведет к увеличению трудоемкости при экономии памяти. Предметом данной статьи является экспериментальное исследование временных характеристик решения задачи коммивояжера методом ветвей и границ с целью определения реального сокращения временных затрат при использовании дополнительной памяти в выбранной структуре хранения дерева решений. Конечной целью исследования является формулировка рекомендаций для реализации метода в практических задачах логистики и бизнес-информатики.

В статье на основе полученных экспериментальных данных показано, что оба рассмотренных варианта классического алгоритма решения задачи коммивояжера методом ветвей и границ порождают программные реализации с экспоненциальной зависимостью времени выполнения от длины входа. Экспериментальные результаты позволяют говорить, что возможность использования дополнительной памяти объемом не более 1 Гб приводит к значительному (до пяти раз) сокращению временных затрат. Прогноз по полученному тренду позволяет сформулировать рекомендацию по практическому применению программной реализации алгоритма метода ветвей и границ с хранением матриц — при реально доступной оперативной памяти в 16 Гб и при ограничении ожидаемого среднего времени счета порядка одной минуты на современных персональных компьютерах могут быть точно решены задачи размерности не более 70.

Литература

1. Little J.D.C., Murty K.G., Sweeney D.W., Karel C. An algorithm for the traveling salesman problem // Operations Research. 1963. No. 11. P. 972—989.
2. Гудман С., Хидетниеми С. Введение в разработку и анализ алгоритмов. М.: Мир, 1981. 368 с.
3. Ульянов М.В. Ресурсно-эффективные компьютерные алгоритмы. Разработка и анализ. М.: ФИЗМАТЛИТ, 2008. 304 с.
4. Ермошин А.С., Плиско В.А. Исследование дерева решений метода ветвей и границ в задаче коммивояжера // Программное и информационное обеспечение систем различного назначения на базе персональных ЭВМ: Межвузовский сборник научных трудов / Под ред. д.т.н., проф. Б.М.Михайлова. М.: МГАПИ, 2006. Вып. 9. С. 76—82.
5. Сигал И.Х., Иванова А.П. Введение в прикладное дискретное программирование: модели и вычислительные алгоритмы. М.: ФИЗМАТЛИТ, 2007. 304 с.
6. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. 2-е издание: Пер. с англ. М.: Вильямс, 2005. 1296 с.
7. Вирт Н. Алгоритмы и структуры данных. Новая версия для Оберона / Пер. с англ. М.: ДМК Пресс, 2010. 272 с.

METHODS OF CONFIGURING THE DECISION-MAKING SYSTEM WHEN CHOOSING AND SUPPORTING AN ENTITY DEVELOPMENT STRATEGY

Victor V. MARSHIROV
Associate Professor, Department of Information Systems and Technologies,
National Research University Higher School of Economics
Address: 25/12, Bolshaya Pecherskaya Street, Nizhniy Novgorod, 603155, Russian Federation
E-mail: vmarshirov@hse.ru

Larisa E. MARSHIROVA
Associate Professor, Department of Accounting, Analysis and Audit,
National Research University Higher School of Economics
Address: 25/12, Bolshaya Pecherskaya Street, Nizhniy Novgorod, 603155, Russian Federation
E-mail: lmarshirova@hse.ru

This article is devoted to the problems of formalization of decision-making in the choice of development strategy in an organization and the ways to implement this strategy. A prioritization method is proposed that allows converting qualitative indicators into qualitative variables by means of pairwise comparison of the objects. As opposed to a simple summation of estimates preferences, this computational algorithm allows one to take into account indirect benefits of all the objects under consideration. The approaches to the ranking of experts and challenges an organization faces at various stages of its development are set forth.

The algorithm is validated on the example of a particular company. The estimates of the priorities of experts are provided; the tree of tasks for which the comprehensive priorities are designed (taking into account the importance and relevance of expert tasks for each expert) is constructed; the analysis of the results for different conditions of the external and internal environment of the organization is made. Recommendations are given on the choice of the deviation values for matrix of pairwise comparisons of objects, as well as a reasonable number of iterations of the calculation of the integrated power of these objects.

The practical significance of the work lies in the fact that the proposed algorithm and methodology for ranking of experts, tasks and subtasks may be used to prepare management accounting regulations to improve decision-making methods, taking into account the strategic and tactical objectives of the organization.

Introduction

E ach entity should have an effective strategy for its successful development, i.e. a development plan for an extended time frame. The significance of strategic management increases in the context of tough competition and a complicated economic situation in the country. The management of a company should ensure fast adaptation of the entity to the changing conditions. The strategy of each entity is unique, as it depends on many factors: business goals, attractiveness of goods, works and services offered, the state of market outlets,

the behavior of competitors, suppliers and customers, production engineering, resources availability and more.

Most often the search for and improvement of development options of an enterprise takes place using brainstorming sessions by business owners and the management team of an entity. However, this process is usually the subject of much controversy. It requires a lot of time for coordination and frequently depends on the authority of the leader, who may authoritatively force the remaining participants of the discussion to make a wrong decision that ignores opinions of all sides involved. When implementing a strategy, financial, human and material resources are frequently deficient, because the plan is to solve all the problems facing an entity at once. The complexity of strategy elaboration and ways of implementing it is also complicated by the fact that various factors affecting this choice are of qualitative or probabilistic nature.

This article suggests a method that can be used to formalize the strategy selection of an enterprise development with due consideration to opinions of all persons involved in decision-making, to rank tasks that must be resolved for implementation of strategy, and regularly to correct this process taking into account problems encountered or new realities.

1. Method of prioritization

In order to improve decision-making processes when choosing ways to implement the strategy of an entity, it is suggested that one address the method of ranking the experts involved and tasks to be solved. A method of prioritization [1] is used in this method. Prioritization assumes pair-wise comparison of objects and filling in the pair-wise comparison matrix $M = \|m_{ij}\|$, wherein at the intersection of the row i and the column j the preference estimate of the object i over the object j must be specified. This estimate is determined as follows:

$$m_{ij} = \begin{cases} 1+y, & \text{if the object } i \text{ is more preferable than the object } j; \\ 1, & \text{if the objects } i \text{ and } j \text{ are equivalent}; \\ 1-y, & \text{if the object } j \text{ is more preferable than the object } i, \end{cases}$$

where y is the deviation of elements of the matrix M, which falls within the limits $0 < y < 1$.

Thereafter, for each object i the integrated force of the first order $F_i(1)$ must be calculated by summing up its preference estimates for the corresponding line:

$$F_i(1) = \sum_{j=1}^{n} m_{ij},$$

where n is the number of the objects.

These integrated forces of the first order can be represented as a vector for all the objects being compared

$$F(1) = \left[F_1(1), F_2(1), ..., F_i(1), ..., F_n(1) \right].$$

However, the index of an integrated force of the first order of the object i $F_i(1)$ does not consider the «force» of other objects, being «won» by the object i or «lost» by the object j in the comparison. Therefore, in order to obtain more precise estimates for each object, one determines the integrated force of the second order with consideration of «forces» of all objects being compared:

$$F_i(2) = \sum_{j=1}^{n} m_{ij} F_i(1),$$

where $i = j = \overline{1, n}$, and n is the number of objects.

Further iterations are made in a similar way, and integrated forces of the k order are calculated using the following formula in the matrix form:

$$F(k) = MF(k-1),$$

where $F(0) = (1,1,...1)$.

This computation algorithm, which is unlike the simple summation of preference estimates, provides consideration of side advantages of all other objects being compared (the issue of iterations quantification will be discussed in Section 3).

For convenience of use and opportunity of comparison, integrated forces of objects of the k order are normalized, i.e., are presented within the scale with a fixed sum of estimates equal to one:

$$F_i^{rel}(k) = \frac{F_i(k)}{\sum_{i} F_i(k)},$$

where $F_i^{rel}(k)$ is the normalized integrated force (priority) of the k order of the object i.

$$\text{Since } \sum_{i=1}^{n} F_i^{rel}(k) = 1,$$

the numerical value of priority $F_i^{rel}(k)$ represents the relative degree of preference of the object i over all other objects being compared.

2. Formulation of the task

Let's consider the use of the method of prioritization on the example of one enterprise in the city of Nizhny Novgorod. The enterprise mentioned has been producing tools for already over ten years. During this time, the enterprise was able to conquer a broad market outlet, but in mid-2014 its profits began to decline. Therefore, the owners decided to identify factors of the deteriorating

performance of the company. Five major reasons for loss of work of the entity were identified: errors in the sales policy resulting in the loss of several important customers; growth of overhead expenses; several kinds of products running at a loss; failures in the processing equipment operation resulting in increased down time and repair costs; increased spoilage in production. In other words, the enterprise was in a pre-crisis state.

To prevent further decline in profits, it was necessary to correct the development strategy of the entity. However, discussion of feasible strategies showed that each owner and each top manager had his own preferences in choosing the plan for the way out of the situation. Therefore, in this situation it was decided to use the method of prioritization.

When choosing a development strategy for the entity, the following stages were subsequently implemented:

✦ appointment of experts and their ranking by assigning weighting indexes to each expert;

✦ formation of the tasks and subtasks tree to implement the strategy of the entity;

✦ task ranking by each expert using the method of prioritization;

✦ computation of the integrated priority of each task on the basis of experts' priorities and the task's importance for each expert.

3. Ranking experts

Two owners (denoted S1 and S2) and three chief executives: the director (D), the deputy director (DD) and the financial director (FD) were experts for choosing the organization's strategy and ways of implementing it. The owners compiled a matrix of experts pair-wise comparison for ranking experts $M = \|m_{ij}\|$, where

$$m_{ij} = \begin{cases} 1+y, & \text{if the expert } i \text{ has greater importance} \\ & \text{in decision making than the expert } j; \\ 1, & \text{if the experts } i \text{ and } j \text{ have equal} \\ & \text{importance in decision making;} \\ 1-y, & \text{if the expert } j \text{ has greater importance} \\ & \text{in decision making than the expert } i. \end{cases}$$

It was decided to determine the deviation value y in conformity with *Table 1*.

Let's consider each option in more detail.

The first option is used when the owners consider that the importance of their estimates substantially exceeds the importance of estimates of other experts. In such a case, e.g., the following ratios are composed:

Table 1.

**Deviation value of elements
of the experts' pair-wise comparisons matrix**

Option	Difference in the degree of experts significance in decision-making	Deviation value (y)
1	high	0.9 – 0.95
2	average	0.5
3	low	0.1 – 0.2

S1 = S2 >> D = DD = FD or S1 = S2 >> D > DD = FD or S1 = S2 >> D > DD > FD and others. In other words, knowledge, experience, intuition, willingness to take responsibility in management decision-making of the owners (S1 and S2) is significantly higher than that of the hired personnel (D, DD, FD), since the sign «>>» is present.

The first option should also be used if the owners consider that the importance of estimates of one or more experts of the hired personnel must be considerably increased, i.e. other inequalities are composed, but therein the sign «>>»is still present: D >> S1 = S2 > DD > FD or D = FD >> S1 > S2 > DD or D >> FD > S1= S2 >DD and others.

The specific value of y is set on the basis of practical considerations.

The second option is expedient to use if the owners when estimating the significance of all experts consider that none of them fits in significantly against others. For example: S1 = S2 > D > DD > FD or S1> S2 > D = DD > FD or S1 = S2 > D > FD > DD and others. It is apparent that in these ratios the signs «>>» are missing and the signs «>» prevail, i.e. the difference in the degree of experts' significance in decision-making is average, so in conformity with *Table 1* $y = 0.5$.

The third option is selected in expert ranking by the degree of their significance when the signs «>>» are missing, and the signs «=» prevail. For example: S1 = S2 > D = DD = FD or S1 > S2 = D = DD = FD or S1 = S2 = D = DD > FD and others. In this case, in conformity with *Table 1*, it is proposed to use $y = 0.1 - 0.2$.

Table 2 shows the results of experts' priorities (weights) computation which was made in mid-2014 for the inequality S1 = S2 >> D > DD = FD. In other words, the owners of the entity assumed for themselves the highest priority when choosing a strategy in comparison with hired managers, and assigned to the director of the entity a higher priority than to the deputy director for operations and to the financial director. Therefore, according to *Table 1* $y = 0.9$.

Table 2.

**Results of priorities (weights) computation of five experts who will choose
the development strategy of the entity and ways of implementing it**

$i \backslash j$	S1	S2	D	DD	FD	F(1)	$F^{rel}(1)$	F(2)	$F^{rel}(2)$	F(3)	$F^{rel}(3)$	F(4)	$F^{rel}(4)$
S1	1	1	1.9	1.9	1.9	7.7	0.31	33.64	0.37	113.24	0.36	388.11	0.36
S2	1	1	1.9	1.9	1.9	7.7	0.31	33.64	0.37	113.24	0.36	388.11	0.36
D	0.1	0.1	1	1.9	1.9	5	0.20	10.91	0.12	42.87	0.14	145.69	0.14
DD	0.1	0.1	0.1	1	1	2.3	0.09	6.64	0.07	21.10	0.07	69.13	0.07
FD	0.1	0.1	0.1	1	1	2.3	0.09	6.64	0.07	21.10	0.07	69.13	0.07
Total						25	1.00	91.47	1.00	311.55	1.00	1060.18	1.00

The first six columns in *Table 2* represent the pair-wise comparison matrix of experts in conformity with the basic formula of the method of prioritization. By summing up the preference estimates in each row, one determines the integrated force of the first order for each expert $F(1)$, which is subsequently normalized to obtain estimates of all experts in fractions of one $F^{rel}(1)$. In order to obtain more accurate estimates for each expert, the integrated forces of the second, third and fourth orders $F(2)$, $F(3)$, $F(4)$ are calculated by the described algorithm, and these are also normalized.

It follows from *Table 2* that after the fourth iteration the weight (priority) of the first two experts was 36%, which is more than 2.5 times higher than the priority of the director of the entity and more than 5 times higher than the weight of his deputies.

After the third iteration, the priority values stopped changing, so 3-4 iterations of corresponding computations can be recommended for practical use.

4. Procedures
of expert ranking
of entity growth option

The owners of the company and its top executive management formulated the prime target that they set for themselves: the withdrawal of the enterprise from pre-crisis status to a higher level of development. To achieve this target with due consideration of all expert opinions, five prime tasks that must be solved were identified: cost savings (C1), increased income (C2), search for financial resources (C3), new production output (C4), improvement of the management system (C5).

But before the experts evaluate them, it is necessary to develop the tree of subtasks to be solved for implementation of each direction of the entity development.

Only in this case can the experts be sure that they equally understand the structure of each task, measures for its achievement, possible risks and restrictions. Therefore, each of the five formulated tasks was structured in *Table 3* below, i.e. the subtasks tree of the first and second levels was developed. Estimates of the resource requirements (financial, human, material) were prepared for each subtask but many of those estimations were of probabilistic nature.

At the next stage, each expert composed the pair-wise comparison matrix of tasks from *Table 3* by the algorithm described in section 1. It was decided to determine the deviation value of elements of these matrices in accordance with *Table 4*.

If the expert considers that one or more tasks are considerably more important than the others, he composes, e.g. the inequality: C1 = C2 >> C3 > C4 > C5 or C4 > C2 > C3 >> C1 > C5 or C3 = C5 > C2 > C1 >> C4 or others, wherein at least one sign «>>»is present. On the basis of this inequality, the adjacency matrix will be compiled, wherein $y = 0.9 - 0.95$. If, according to an expert, the tasks have approximately equal significance, then $y = 0.1 - 0.2$. In all other cases, the adjacency matrix will be compiled with the deviation value equal to 0.5. *Table 5* shows the inequalities on the basis of which the experts compiled their adjacency matrices in the middle 2014.

When processing the matrices obtained from the experts, the order equal to four was used. The normalized integrated force of each task was determined by the same formulas that have been described above when calculating the significance of each expert. The computation results are summarized in *Table 6*, and then in the same table all five tasks were ranked taking into account the experts' priorities and the task's importance for each expert.

Table 3.

Tasks and subtasks tree for implementation of the entity's strategy

Tasks	Subtasks of the first level	Subtasks of the second level
C1. Costs saving	Reduction of production cost	Reduction of direct costs (material savings, spoilage reduction, quality control of materials when received from suppliers, improvement of the worker remuneration system)
		Reduction of overhead costs (elimination of redundant functions of management personnel; improvement of the remuneration system, particularly bonuses, appraisal and reduction of production management personnel)
		Introduction of efficient management accounting of expenses and net cost of finished production
		Transfer of cost-based production functions to outsourcing
	Reduction of unprofitable production	Refusal to carry loss-making assortment of finished production
		Certification and reduction of production personnel
		Lease of vacated premises and equipment
		Sale of unnecessary equipment
C2. Increase of income	Increase in sales revenue offinished products	Search for new customers (marketing development, flexible discount system, participation in tenders to purchase tools for state needs)
		Price increase for some types of products (sales analysis, analysis of current market prices in the region)
	Increase in other sales	Sale of unused materials (inventory of stocks, search for buyers)
		Sale of unused fixed assets (preparation for sale, substantiation of selling price, search for buyers)
C3. Search for financial resources	Receiving favorable financing	Bank credits (monitoring terms and interest rates, developing competent business plans)
		Loans (from partner companies, private individuals including the owners)
		Commercial credits obtaining from suppliers and contractors
	Reduction of receivables	Ongoing analysis of accounts receivable and work to prevent overdue debts
		Law suits against debtors related to overdue account receivables
		Sale of bad debts to collection agencies
C4. New production output	Selection of a new range of products	Complex of research and development activities
	Preparation for serial production of new products	Purchase and installation of equipment, new materials and components
		Development of standards and regulations
		Personnel training
		Advertising
	Search for new contractual partners	Search for suppliers, contractors (analysis of prices, quality of supplied materials, works and services)
		Search for buyers (analysis of prices, flexible discount system, advertising)
C5. Improvement of management system	Introduction of budgeting	Development of the financial model of entity, allocation of financial responsibility centers
		Development of planning regulations, accounting, analysis and plans revision
	Changing the employees' incentive system	Development of new regulations for material incentives of employees
		Development of new regulations for non-financial recognition of employees
	Reduction of administrative staff	Regular appraisal of the managerial staff
		Substantiation of the number of managerial staff and reduction of inefficient managers
	Reduction of commercial and administrative expenses	Analysis of transport, business travel and entertainment expenses; expenses for advertising, office and warehouse space maintenance, and their optimization
		Transfer of cost-based management functions to outsourcing

Table 4.

Deviation value of elements of the pair-wise comparison matrix

Option	Difference in the degree of task importance for a particular expert	Deviation value (y)
1	high	0.9 – 0.95
2	average	0.5
3	low	0.1 – 0.2

Table 5.

Tasks ranking by experts and choosing the deviation value of the adjacency matrix

Experts	Tasks ranking	Deviation value of the adjacency matrix (y)
The first owner (S1)	C4 >> C2 > C1 = C3 = C5	0.9
The second owner (S2)	C1 = C2 = C4 > C3 = C5	0.2
Director (D)	C1 = C2 = C5 > C3 > C4	0.5
Deputy director (DD)	C4 >> C1 = C5 > C2 > C3	0.9
Financial Director (FD)	C1 > C2 > C3 = C4 > C5	0.5

Table 6.

Results of tasks ranking by each expert and the integrated priority of each task

Task	Experts										Priority	Place
	A1		A2		D		DD		FD			
	0.37	place	0.37	place	0.14	place	0.07	place	0.07	place		
C1	0.11	3-5	0.22	1-3	0.24	1 -3	0.21	2-3	0.29	1	0.19	3
C2	0.22	2	0.22	1-3	0.24	1 -3	0.07	4	0.24	2	0.21	2
C3	0.11	3-5	0.18	4-5	0.15	4	0.04	5	0.17	3-4	0.14	5
C4	0.44	1	0.22	1-3	0.12	5	0.47	1	0.17	3-4	0.30	1
C5	0.11	3-5	0.18	4-5	0.24	1 -3	0.21	2-3	0.12	5	0.16	4
Total	1.00		1.00		1.00		1.00		1.00		1.00	

It is apparent that the following information is contained in *Table 6* for each expert:

✦ the priority that he got at the previous stage of computation (see *Table 2*), e.g., for owners (A1 and A2) — 0.37;

✦ task priorities (C1 − C5), which have been computed on the basis of the pair-wise comparison matrix, e.g., for the general director (D) it is 0.24, 0.24, 0.15, 0.12 and 0.24;

✦ the place of each task in the ranking composed by an expert, e.g., the financial director (FD) in the first place assigned the task of costs saving (C1), in the second − increase of income (C2), in the last place − improvement of the management system (C5), while the tasks related to search for financial resources (C3) and the need for new production output (C4) − share the third and the fourth places.

The penultimate column of *Table 6* contains the integrated priority of each task which has been calculated taking into account the significance of experts and the importance of this task for each expert. It is apparent that the main task for the implementation of the company's strategy is C4 (the need for new production output), which is more than two times more important than task C3 (search for financial resources), which has the lowest priority.

Clearly, for the withdrawal of the enterprise from pre-crisis status it is necessary to solve all five formulated tasks, however, the lack of material, labor, financial and timing resources requires determination of the direction of the «main effort», of that task, which has bigger importance. Each time, making managerial decisions on the distribution of limited resources when developing an action plan for a quarter, month, week, the organization's chief executives and owners had exact priorities for the choice stated in *Table 6*.

In early November 2014, the enterprise started production of new types of products, and several important subtasks in other directions were successfully implemented as well. For this reason, a revision of task priorities was required. This time, the owners decided to raise the significance of the experts − heads of the entity, and the inequality S1 = S2> D = DD = FD was used. In addition, amendments and changes were made in the tasks and subtasks tree, although the task list remained unchanged. Using the algorithms described above, the tasks of costs saving (C1) and improvement of management system (C5) were the first and the second in the ranking.

The entity has gradually accumulated positive experience of applying the method of prioritization. A clear algorithm and the simplicity of the method implemen-

tation made it possible to develop corporate regulations and to implement ranking not only of tasks, but also of subtasks of the first level, involving heads of departments on an expert basis. The time was reduced for discussion and choosing managerial solutions; the possibility appeared to compare not only quantitative, but also qualitative indicators. The profit of the entity in the second half of 2014 as compared to the first half of the same year increased by 28%. The main target of withdrawing the enterprise from the pre-crisis state was achieved.

At the end of 2014 due to changes in the economic situation in the country, the entity once again had to revise the task priorities. This time the need to search for financial resources was recognized as the most important task, while the task of releasing new types of products was excluded from discussion. Financial and legal departments developed a detailed subtask tree in order to obtain additional sources of funding, and subtasks of the first and second levels were ranked using the method of prioritization.

The authors of this article did not have as their goal to improve the Saaty method [2], wherein a rating scale from 1 to 9 is used for objects comparison. These estimates have the following values: 1 — equal importance, 3 — moderate superiority of one above the other, 5 — substantial superiority of one above the other, 7 — drastic superiority of one above the other, 9 — very strong superiority of one above the other; 2, 4, 6, 8 — corresponding intermediate values. If during the pair-wise comparison, an expert considers that the object i considerably exceeds the object j, then in the pair-wise comparison matrix $B = \|b_{ij}\|$ $b_{ij} = 3$, and $b_{ji} = 1/3$. The method of prioritization is just one of possible modifications of the Saaty method.

The authors believe that for an expert, when comparing objects it is easier to answer «better (more important, preferred)», «worse (less important or less preferred)», «equal», than using the scale from 1 to 9 to assign instead the grade 6, e.g. the grade 8 and to be wrong. For this reason, the ranges of parameter y proposed in *Tables 1 and 4* must consider that the spread of estimates can be evaluated prior to the objects comparison. Certainly, the ranges of parameter y and the number of these ranges are only recommended, and thereafter it is possible to correct them depending on the current situation and on the standing tasks.

Conclusion

Consequently, the authors have obtained the following results:

✧ a new approach has been suggested to the procedure of choosing an entity's development strategy and ways of implementing this strategy;

✧ the algorithm of the method of prioritization has been developed, making it possible to compare objects by qualitative factors and obtain quantitative estimates;

✧ procedures for ranking experts and tasks that an entity faces at various stages of its development have been proposed;

✧ the validation of the algorithm has been implemented using a particular enterprise: the estimates of experts' priorities have been presented; a task tree has been composed for which integrated priorities have been computed with due consideration of experts' significance; an analysis has been carried out of the results obtained for different conditions of external and internal environment of an entity.

The practical significance of the work lies in the fact that the proposed algorithm and the method of ranking experts and tasks can be used to prepare management accounting regulations to improve decision making methods with due account of the strategic and tactical tasks of an entity. ∎

References

1. Marshirov V.V., Marshirova L.E. (2013) Kompleksnaya otsenka individual'nogo truda razrabotchikov programmnogo obespecheniya [Comprehensive evaluation of individual work of software developers]. *Business Informatics*, no. 2 (24), pp. 55–62 (in Russian).
2. Saaty T.L. (2001) *Decision making for leaders: The Analytic Hierarchy Process for decisions in a complex world.* Pittsburgh, PA: RWS Publications.

МЕТОДИКА ОРГАНИЗАЦИИ СИСТЕМЫ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВЫБОРЕ И ПОДДЕРЖКЕ СТРАТЕГИИ РАЗВИТИЯ ОРГАНИЗАЦИИ

В.В. МАРШИРОВ
кандидат технических наук,
доцент кафедры информационных систем и технологий,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 603155, г. Нижний Новгород, ул. Большая Печерская, 25/12
E-mail: vmarshirov@hse.ru

Л.Е. МАРШИРОВА
кандидат экономических наук, доцент кафедры бухгалтерского учета,
анализа и аудита, Национальный исследовательский университет
«Высшая школа экономики»
Адрес: 603155, г. Нижний Новгород, ул. Большая Печерская, 25/12
E-mail: lmarshirova@hse.ru

Статья посвящена проблемам формализации процессов принятия решений при выборе стратегии развития организации и путей реализации этой стратегии. Предложен метод расстановки приоритетов, который позволяет путем попарного сравнения объектов переводить качественные показатели в количественные. Этот алгоритм расчета, в отличие от простого суммирования оценок предпочтения, позволяет учесть косвенные преимущества всех других объектов, с которыми производилось сравнение. Изложены подходы к ранжированию экспертов и задач, которые стоят перед организацией на разных этапах ее развития.

Проведена апробация алгоритма на примере конкретного предприятия: представлены расчеты приоритетов экспертов, построено дерево задач, для которых рассчитаны комплексные приоритеты с учетом значимости экспертов и значимости задач для каждого эксперта, а также проведен анализ полученных результатов для различных условий внешней и внутренней среды организации. Даны рекомендации по выбору значений величины отклонений матриц парных сравнений объектов и обосновано количество итераций при расчете интегрированной силы этих объектов.

Практическая значимость работы заключается в том, что предложенные алгоритм и методика ранжирования экспертов, задач и подзадач могут быть использованы при подготовке регламентов управленческого учета для совершенствования методов принятия решений с учетом стратегических и тактических задач организации.

Литература

1. Марширов В.В., Марширова Л.Е. Комплексная оценка индивидуального труда разработчиков программного обеспечения // Бизнес-информатика. 2013. № 2 (24). С. 55–62.
2. Saaty T.L. Decision making for leaders: The Analytic Hierarchy Process for decisions in a complex world. Pittsburgh, PA: RWS Publications, 2001. 343 p.

MATHEMATICAL MODEL AND ALGORITHM OF SELECTING SOFTWARE PROMOTION OPTIONS DIFFERENTIATED BY FUNCTIONALITY AND BUSINESS MODELS

Yuriy P. YEKHLAKOV
Professor, Head of Department of Data Processing Automation,
Tomsk State University of Control Systems and Radioelectronics (TUSUR)
Address: 40, Lenina ave., Tomsk, 634050, Russian Federation
E-mail: upe@tusur.ru

Dmitriy N. BARAKSANOV
Post-graduate Student, Department of Data Processing Automation,
Tomsk State University of Control Systems and Radioelectronics (TUSUR)
Address: 40, Lenina ave., Tomsk, 634050, Russian Federation
E-mail: upe@tusur.ru

The paper focuses on issues and tasks IT companies face with when promoting proprietary software products on consumer markets. It provides the description of the basic software product market represented as a combination of specific prospective customer groups, to which a software product line with relevant functionality can be offered using different delivery business models. The authors suggest three groups of market segment attractiveness: market attractiveness, competition level and a company's performance in the segment.

The authors developed a multi-objective mathematical model for selecting options of software product distribution in selected market segments. The following selection criteria for software distribution options in selected market segments are advanced: maximum of total profit, minimum of promotion expenses, minimum of target segments number and maximum of an integrated measure of segments attractiveness. Limitations of the model include the volume of human resources in each single-discipline group of specialists in an IT company and application of a single option of software product delivery for each of market segments. An algorithm for solution of the problem relies on successive concession method. However, suggested criteria of optimization are applicable for other methods of the multi-objective problem solution.

The authors described the results of the model application for a real problem of selecting distribution options for the 'Electronic Timetable' software product for specialized secondary education and higher education institutions of Siberian Federal District in Russia. Twenty-four segments were selected for distribution, where two software distribution models were possible — «software as a service» (SaaS) and «application service provider» (ASP). Assessment of qualitative and quantitative parameters of the mathematical model was conducted by experts using optimistic, pessimistic and realistic scenarios. As a result of the problem solution, software product should be distributed as SaaS in 14 segments, and as ASP — in 3 segments. The results of the study may be useful for directors and managers of small IT companies when formulating strategies for proprietary software product distribution in certain target segments.

Introduction

At present, IT companies associated with production of software products, in one way or another select one of the two activity business models: development and promotion of their own software products (product model), or development of a unique customized software product (custom model). From the developer's point of view, the product model is more promising by virtue of the fact that the company is positioned on the market as a producer of new projects and technologies. In this context, the company management is to solve such tasks as determining the boundaries of the primary software market (distinguishing target segments therein), building functionally different options of software products architectures, identifying business models of their delivery to the consumers, as well as selecting the optimal range of software delivery options for target market segments in the resource-constrained environment.

From the point of view of marketing and management [1, 2], these tasks can be resolved by determining a product-market matrix in the form of a variety of product offerings that meet the consumers' requirements in selected target market segments and developing a matrix-based product portfolio, providing a required balance between the producer resource capabilities, consumer needs, amount of a potential gain and expected cost. In publications related to management of IT companies, the tasks of this sort are assigned to portfolio management. Papers [3—5] describe capabilities of specific market tools to manage the project portfolios that provide information support for the basic software projects life cycle processes. Description of models and procedures to optimize software projects management is given in [6, 7]. Furthermore, the literature currently discusses an area of engineering lines of software products. For example, papers [8—10] present a policy of industrial software development, based on organization of a family (line) of software products by reusing components. Paper [11] addresses a two-stage model of generating a product line satisfying the preferences of a wide range of consumers and developing an optimal solution for release of products in demand by a certain group of consumers.

With due regard of the above approaches, this article addresses the task of launching replicated software products different in the functionality and business models of delivery to the target markets in the IT company resource-constrained environment.

1. Description of the primary market

The three-dimensional scheme proposed by D.Abel and improved by J.-J. Lambin [2] was taken as a basis of determining the software product primary market boundaries, where groups of potential customers, functions (consumer needs) and technologies (alternative ways of implementing the functions) were identified as the original coordinates determining the primary market. To determine the boundaries of the primary market of replicated software, it is proposed to use the following coordinate system: target market segments on which the company will concentrate its efforts differentiated with respect to the functionality of software delivery options; software delivery business model and associated services (*Fig. 1*).
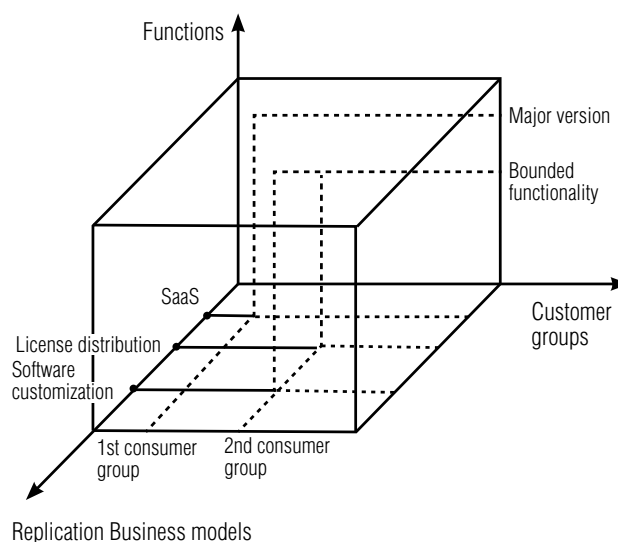
Fig. 1. 3D model describing software market

For selection of target market segments and groups of potential customers both hierarchical and iterative segmentation methods [12] can be used. In so doing, in the light of specific features of software as a commodity good, the following set of characteristics as segmentation variables is proposed to be used: geographical location, type of ownership, sectoral affiliation, corporate consumer size, a list of purchased services, level of development of information and communication technologies [13].

The needs of potential consumers in each of the target segments can be satisfied through delivery of functionally differentiated software versions: major version or differentiated set of functions (depending on software configuration or integrity of its delivery).

Free software distribution with further maintenance; distribution of licenses for box versions; Software as a Service, (SaaS); dedicated installation of the software (Application Service Provider, ASP); software customization as required by the customer can be distinguished as the main software delivery business models and associated services.

In consideration of the foregoing, a description of the software primary market can be represented as a set S of certain groups of potential customers which software with functionality F of interest based on a specific software delivery business model B can be proposed to.

2. Mathematical model and solution algorithm

Let us assume that $S = \{1, 2, ..., j, ..., m\}$ is a set of target market segments resulting from segmentation of consumers of the primary market; $F = \{1, 2, ..., f, ..., d\}$ is a set of functionally differentiated software versions; $B = \{1, 2, ..., b, ..., l\}$ is a set of software delivery business models and associated services for each of target market segments. Software delivery of a specific functionality replicated by a specific business model including a number of service applications, such as server-based version, desktop-applications, mobile applications, etc. will be identified as an alternative software delivery. The Cartesian product of the sets $F*B$ forms a set of alternatives (options) of the software product delivery to the target market segments $V = \{1, 2, ..., i, ..., n\}$. Then, the mathematical model of selecting software delivery options for the target market segments can be represented as follows.

Set $X = \{x_{ij}\}$, $i = \overline{1,n}$, $j = \overline{1,m}$, shall be determined, where

$$x_{ij} = \begin{cases} 1, \text{ if the } i\text{-th delivery option will progress} \\ \quad \text{in the } j\text{-th target segment;} \\ 0, \text{ if this is not the case.} \end{cases}$$

It is obvious that the selection of a specific option of software product delivery depends on the company's abilities to attract single-discipline specialists (programmers, system administrators, sales managers, support professionals, and others.) for replication processes. In this case, the restriction of a task on human resources can be represented as

$$\sum_{i=1}^{n}\sum_{j=1}^{m} t_{ijk} \cdot x_{ij} \leq T_k, \ k = 1,t \qquad (1)$$

where t_{ijk} are required human resources (in man-hours and man-months) of the k-th group of single-discipline specialists to ensure replication processes of the i-th

software delivery option in j-th segment;

T_k are human resources available at the company of the k-th group of single-discipline specialists.

A small IT company is interested in replication of only one software delivery option to each segment:

$$\sum_{i=1}^{n} x_{ij} = 1, \ j = \overline{1,m}. \qquad (2)$$

This can enhance the quality of the product positioning in the target market segment, ensure specialization of profession-oriented corporate employees for the segment customer service and preclude the need for supporting several software delivery options. As a final result, this leads to a decrease in an overall effort to ensure the software replication processes.

It is customary that two indicators affecting the efficiency of business operations are applied as an optimality criterion in the problems of this kind:

1) maximization of accumulated profit in software replication:

$$Z_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \cdot x_{ij} \rightarrow max, \qquad (3)$$

where P_{ij} is standard (desirable) profit from the i-th option of software delivery to j-th segment;

2) minimization of expenses (investments) for software promotion to target markets:

$$Z_2 = \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij} \cdot x_{ij} \rightarrow min, \qquad (4)$$

where v_{ij} is a planned value of investments for promoting the i-th option of software delivery in j-th segment.

Furthermore, IT companies, due to limited financial and human resources, are unable to meet the requirements of the whole market, so the management task is to select a minimum number of target segments, which servicing will enable to reduce the potential cost of the product promotion:

$$Z_3 = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \rightarrow min, \qquad (5)$$

The classical marketing literature [1, 2] points out that in order to estimate the efficiency of business operations in a particular market segment, it is reasonable to consider the qualitative characteristics of segments attraction along with the evaluation of quantitative indicators. With this in mind, an integrated indicator of the attraction of business operations as a target function can be used

$$Z_4 = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} \cdot x_{ij} \rightarrow max, \qquad (6)$$

where w_{ij} is an integrated attraction of replication of the *i*-th option of software delivery in *j*-th segment.

Mathematical model (1-6) is a multi-criterion problem of the linear integer programming. Generally speaking, the dimension of problem is determined based on the amount allocated segments and proposed for replication of software delivery options.

With due regard to a high uncertainty of the initial model parameters and ambiguity of the relative importance of criteria, a method of successive concessions [14] for solution of the problem will be used, according to which local criteria are pre-ranked by a decision maker (DM) in a descending order of their relative importance. In this case, criterion Z_1, is considered to be the most important, and criteria — Z_2, Z_3 and Z_4 are considered to be less important.

In the first step, using expression (3) as an optimality criterion, we solve a one-criterion problem of linear programming (1-3). The resulting optimal solution of the problem (1-3) makes it possible to determine options of software delivery to specified segments with the value of expected profit *P* In the second step the decision-maker assigns a value of allowable reduction of profit $\Delta P \geq 0$. Expression (4) is used as the optimality criterion, and expression (3) is transferred to the category of restrictions (7). A solution of one-criterion problem of linear programming (1, 2, 4, 7) is found:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \cdot x_{ij} \geq P - \Delta P. \qquad (7)$$

Concession $\Delta V \geq 0$ is assigned for second criterion Z_2, expression (4) is transferred to a category of restrictions (8). The solution of one-criterion problem of linear programming (1, 2, 5, 7, 8) for criterion Z_3 is found with specified concessions for the first and second criteria:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij} \cdot x_{ij} \leq V + \Delta V, \qquad (8)$$

Further, concession $\Delta S \geq 0$ is made, expression (5) is transferred to a category of restrictions (9) and problem (1, 2, 6-9) is solved using criterion Z_4. The resulting candidate solution is considered to be optimal.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \leq S + \Delta S. \qquad (9)$$

3. Selection and justification of the list and evaluation of the primary model indicators

Based on the analysis and generalization of papers [15-17], the list of primary indicators for calculation of

an integrated index for assessing the business operations appeal is proposed to be determined proceeding from the market appeal of target segments, software competitiveness in the target segments and capabilities of the company activity in these segments (T*able 1*).

Under conditions of a high degree of uncertainty for estimation of quantitative parameters of the mathematical model and qualitative indicators of the segment appeal an expert evaluation method in conjunction with PERT method [18] is proposed to be used. The evaluations of target segments appeal indicators can be determined by the experts based on optimistic o, pessimistic p and realistic b options (*Table 2*). An average evaluation of the appeal is determined by $w_{ij}^{k} = (w_{ij}^{o} + 4w_{ij}^{b} + w_{ij}^{p})/6$.

Table 1.

Indicators of market appeal of target segments

Indicator	Indicator description
Indicators of market appeal of target segments	
Segment capacity and trend	Status of growth dynamics or reduction of a number of enterprises being potential consumers of software
Information infrastructure level	Availability and quality of information services at the segment enterprises
Expandability of a range of software and services	Potential demands of enterprises in other company products and services
Indicators of software competitiveness	
Compliance of software characteristics with consumer demands	The extent of required modifications (adaptation) of software to satisfy demands of the segment consumers
Presence of software competitors and brand awareness	Presence and quantity of software competitors similar in performance
Level of software uniqueness	Available of unique advantages in software over similar products.
Indicators reflecting capability of company best practice in a segment	
Consumers availability	Company's ability to conduct necessary communications with the segment consumers, both in terms of software delivery and subsequent maintenance
Technological mobility of the company	Capability and speed of company response to appearance of new segment players and software
Company experience in consumer servicing	The success story of the company cooperation with enterprises similar to the segment consumers by form of ownership, company size, software implementation and maintenance technology, procurement specifics and other characteristics
Compliance with strategic objectives and key competence of the company	The degree of the company conformity with business profile of the segment consumers

The optimistic appeal assessment assumes that the segment is extremely attractive for software replication, the software product is unique, the company has highly qualified specialists, the business profitability is quite high, and occurrence of any risks is improbable. The pessimistic appeal assessment is performed if the company has many problems in the target segment, namely, presence of competitors, high cost, low level and quality of information services at the enterprises and others. A realistic assessment means that the segment is moderately attractive for software replication.

A discussion as an open panel discussion of individual appeal characteristics with the involvement of relevant company specialists and external experts (marketing specialists, analysts, investors, etc.) is proposed to be used as an expert assessment procedure. The project manager can act as a leading discussion manager. Final optimistic, most likely and pessimistic assessments are documented after reaching an opinion consistency of experts involved in the assessment.

The quantitative model parameters, profit thresholds and necessary expenditure are assessed by the same procedure, in so doing, the expert assessments consist of three numbers: minimum, maximum, and most likely values.

4. Test case

Let us consider the process of solving the problem of selecting options of the software product delivery «Electronic timetable» to the market of secondary and higher educational institutions of the Siberian Federal District (SFD) of Russia. The subjects forming part of SFD; types of educational institutions (university or college); size of the institution (number of study groups of students) were used as segmentation variables in differentiation of consumer groups. These segments characteristics were formed on the basis of reference book of federal portal «Russian Education» as of 13.01.2015 [19].

As a result of the market segmentation 24 target segments of special secondary and higher educational institutions of the Siberian Federal District (12 subjects by 2 types of educational institutions with not more than 100 study groups) were obtained. «Electronic timetable» software is planned to be replicated in two delivery business models: SaaS model with a built-in mechanism of adaptation enabling the consumers to work with a single cloud service, and ASP model representing an allocation of a standalone application for a separate educational institution with more powerful capabilities of adaptation and improvement according to the customer needs. The SaaS model can be used by colleges and universities, while the ASP model can be used by institutions only. *Table 2* presents a fragment of initial data for solving the problem of selecting «Electronic timetable» software delivery options to SFD universities and colleges.

The results of the iterative solution of the problem based on the method of successive concessions are presented in *Table 3*. A set of «Electronic timetable» software replication options resulting from the last iteration looks like this:

1) **delivery of SaaS-version:** colleges and universities of the Altai Territory; colleges of Zabaikalye Territory; colleges and higher educational institutions of Irkutsk region; colleges and universities of Kemerovo region; colleges and universities of Krasnoyarsk Territory; colleges and universities of Novosibirsk Region; colleges of Omsk region; colleges of the Republic of Buryatia; colleges Tomsk region.

2) **delivery of ASP-version:** universities of Omsk region; universities of the Republic of Buryatia; universities of Tomsk region.

Table 2.

**Characteristics of options of «Electronic timetable»
software delivery to universities and colleges of the Siberian Federal District**

Subject of SFD	Appeal			Profit, thousand rubles			Expenses for promotion thousand rubles			Technicians, man-hours		Project Manager, man-hours			
	1	2	3	1	2	3	1	2	3	1, 2	3	1	2	1, 2	3
Altai territory	7,9	8	8,5	78,4	29,4	47,6	33,6	12,6	20,4		500	40	15		124
Novosibirsk region	8	8,1	8,6	98	29,4	47,6	42	12,6	20,4	300	500	50	15	96	124
⋮															
Tomsk region	7,8	8,2	8,7	39,2	19,6	23,8	16,8	8,4	10,2		250	20	10		62

Delivery options: 1 – delivery to colleges using SaaS model;
2 – delivery to universities using SaaS models; 3 – delivery to universities using ASP model

Table 3.

**Results of incremental solution of problems
for selecting «Electronic timetable» software delivery options**

Criteria	INDICATORS					
	Number of delivery options	Average appeal	Expected profit, thousand rubles	Expected expenses for promotion, thousand rubles	Required human resources of technicians, man hour	Required human resources of project managers, man hour
1. Profit maximization	22	7,79	559,9	417	1050	742
2. Minimization of expenses for promotion	20	7,74	501,1	391,8	1050	712
3. Minimization of a number of target segments	17	7,8	501,1	391,8	1050	712
4. Maximization of average appeal	17	7,87	505,3	393,6	1300	764

The analysis of the incremental selection of «Electronic timetable» software delivery options enables to conclude that a number of selected segments (from 22 to 17) is changed in various iterations of solving the problem. The results obtained in the second and third steps show that while minimizing a number of segments the expected profit and expenses remain unchanged, and the amount of human resources increases. This may be due to the fact that at these algorithm stages segments with a large number of educational institutions are selected. A decrease of the expected profit to expense ratio (from 1.34 to 1.28) is observed between the first and the second steps, however, in the next steps this ratio is stabilized, at this, a steady growth of the average segments appeal indicator is observed from the third step.

Conclusion

A description of the primary software market in the form of a three-dimensional model (consumers, functionality, delivery business model) enables to create a base set of software delivery options and prioritize delivery options based on the company's own capabilities and specifics of the market of potential users. The four optimality criteria proposed in the article make it possible to describe the problem of selecting functionality- and business model-differentiated options for software delivery in the form of alternative models, and use other methods of solving multi-criteria problems. In addition, the proposed model of volume planning of selection of software delivery options to the target market segments can be modernized in a volume-scheduling model that accounts for software time intervals of delivery (for example, year, quarter, month).

The analysis of the results of selecting options of the «Electronic timetable» software delivery to the market of specialized secondary and higher educational institutions of the Siberian Federal District makes it possible to conclude about the suitability of the proposed model and algorithm for solving practical problems. The problem solving results may be useful to directors and managers of small IT companies in forming replication strategies for their own software products to prospective target segments. ■

References

1. Webster F. (2005) *Osnovy promyshlennogo marketinga* [Fundamentals of industrial marketing]. Moscow: Grebennikov Publishing House (in Russian).
2. Lambin J.-J. (2007) *Menedzhment, orientirovannyj na rynok. Strategicheskij i operacionnyj marketing* [Market oriented management. Strategic and operational marketing]. St.Petersburg: Piter (in Russian).
3. Chulani S., Santhanam P., Hodges B., Anders K. (2007) Metriki upravlenija portfelem programmnyh produktov [Metrics-based management of software product portfolios]. *Open Systems*, no. 3. Available at: http://www.osp.ru/os/2007/03/4177900/ (accessed 9 February 2012) (in Russian).
4. Dubova N. (2008) Upravlenie s portfelem [Management with portfolio]. *Open Systems*, no. 3. Available at: http://www.osp.ru/os/2008/03/5015107/ (accessed 9 February 2012) (in Russian).
5. Oganesyan A. (2008) Opyt upravlenija portfelem IT-proektov [Experience of managing a portfolio of IT projects]. *Open Systems*, no. 3. Available at: http://www.osp.ru/os/2008/03/5015181/ (accessed 9 February 2012) (in Russian).
6. Avdoshin S.M., Lifshits A.A. (2014) Formirovanie portfelja proektov na osnove nechetkoj modeli mnogokriterial'noj optimizacii [Project portfolio formation based on fuzzy multi-objective model]. *Business Informatics*, no. 1 (27), pp. 14–22 (in Russian).

7. Nedovesov M.V., Rudenko Z.G. (2012) Formirovanie optimal'nogo portfelja vzaimozavisimyh proektov i ego optimizacija po vremeni [Optimal portfolio of interdependent projects construction and time optimization]. *Control Sciences*, no. 4, pp. 26–31 (in Russian).

8. Kang K., Lee J., Donohoe P. (2002) Feature-oriented product line engineering. *IEEE Software*, July/August, pp. 58–65.

9. Koznov D.V., Novickij I.A., Smirnov M.N. (2013) Instrumenty dlja upravlenija variantivnost'ju: gotovnost' k promyshlennomu primeneniju [Variability management tools: Readiness for industrial use]. *SPIIRAS Proceedings*, no. 3 (26), pp. 297–331 (in Russian).

10. Guss S.V. (2011) Razrabotka semejstva programmnyh sistem v specificheskoj predmetnoj oblasti [The development of a family of software systems in a particular subject area]. *Mathematical Structures and Modeling*, no. 22, pp. 55–68 (in Russian).

11. Michalek J.J., Ebbes P., at all. (2011) Enhancing marketing with engineering: Optimal product line design for heterogeneous markets. *International Journal of Research in Marketing*, no. 28, pp. 1–12.

12. Kosterin A.G. (2002) *Praktika segmentirovanija rynka* [The practice of market segmentation]. St.Petersburg: Piter (in Russian).

13. Yekhlakov Yu.P., Baraksanov D.N., Mamonova N.V. (2012) Funkcional'naja i matematicheskie modeli segmentirovanija rynka programmnyh produktov [Functional and mathematical models of segmentation consumers of software]. *St.Petersburg State Polytechnical University Journal. Computer Science. Telecommunications. Control Systems*, no. 2, pp. 155–160 (in Russian).

14. Podinovsky V.V., Gavrilov V.M. (1975) *Optimizacija po posledovatel'no primenjaemym kriterijam* [Optimization on consistently applied criteria]. Moscow: Sovetskoe radio (in Russian).

15. Wood M.B. (2007) *Marketingovyj plan. Prakticheskoe rukovodstvo po razrabotke* [The marketing plan: A handbook]. Moscow: Williams (in Russian).

16. *Analiz assortimenta s pomoshh'ju matricy McKinsey – General Electric* [Analysis of the range using the McKinsey – General Electric matrix]. Available at: http://powerbranding.ru/biznes-analiz/matrica-mckinsey/ (accessed 20 June 2014) (in Russian).

17. *Chto takoe privlekatel'nost' rynka i kak ee ocenit'?* [What is the attractiveness of the market and how to evaluate it?]. Available at: http://powerbranding.ru/rynok/ocenka-privlekatelnosti/ (accessed 20 June 2014) (in Russian).

18. Fatrell R.T., Shafer D.F., Shafer L.I. (2004) *Upravlenie programmnymi proektami. Dostizhenie optimal'nogo kachestva pri minimume zatrat* [Software project management. Achieving optimal quality at the lowest cost]. Moscow: Williams (in Russian).

19. *Rossijskoe obrazovanie: Spravochnik* [Russian education: A reference book]. Available at: http://www.edu.ru/ (accessed 09 February 2015) (in Russian).

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И АЛГОРИТМ ВЫБОРА ДИФФЕРЕНЦИРОВАННЫХ ПО ФУНКЦИОНАЛУ И БИЗНЕС-МОДЕЛЯМ ПОСТАВКИ ВАРИАНТОВ ПРОГРАММНОГО ПРОДУКТА

Ю.П. ЕХЛАКОВ
доктор технических наук, профессор,
заведующий кафедрой автоматизации обработки информации,
Томский государственный университет систем управления и радиоэлектроники (ТУСУР)
Адрес: 634050, г. Томск, пр. Ленина, д. 40
E-mail: upe@tusur.ru

Д.Н. БАРАКСАНОВ
аспирант кафедры автоматизации обработки информации,
Томский государственный университет систем управления и радиоэлектроники (ТУСУР)
Адрес: 634050, г. Томск, пр. Ленина, д. 40
E-mail: bdn@tusur.ru

В статье рассматриваются проблемы и задачи, с которыми сталкиваются ИТ-компании при продвижении на рынок собственных разработок. Представлено описание базового рынка программных продуктов (ПП) в виде совокупности определенных групп потенциальных потребителей, которым можно предложить линейку ПП с интересующим их функционалом по различным бизнес-моделям поставки. Предложены три группы показателей привлекательности сегментов рынка – рыночной привлекательности, уровня конкуренции, эффективности работы компании в сегменте.

Разработана многокритериальная математическая модель выбора вариантов тиражирования ПП в выделенных сегментах рынка. В качестве критериев оптимальности выделены максимум суммарной прибыли, минимум затрат на продвижение, минимум количества целевых сегментов и максимум интегрального показателя привлекательности сегментов. Ограничения модели – объемы трудовых ресурсов по каждой группе узкопрофильных специалистов ИТ-компании и поставка в каждый из сегментов только одного варианта ПП. В качестве алгоритма решения задачи применяется метод последовательных уступок, однако предложенные критерии оптимальности позволяют использовать и другие методы решения многокритериальной задачи.

Описаны результаты использования модели на примере реальной задачи выбора вариантов поставки ПП «Электронное расписание занятий» на рынок средних специальных и высших учебных заведений Сибирского федерального округа Российской Федерации. Для тиражирования выделены 24 сегмента, потребителям которых предлагается поставка ПП по двум бизнес-моделям: «software as a service» (SaaS) и «application service provider» (ASP). Оценка количественных и качественных параметров математической модели проведена экспертами по оптимистическому, пессимистическому и реалистическому сценариям. Результаты решения задачи: ПП по модели SaaS предлагается тиражировать в 14 сегментах, по модели ASP – в 3 сегментах. Полученные результаты могут быть полезны руководителям и менеджерам небольших ИТ-компаний при формировании стратегий тиражирования собственных программных продуктов в перспективных целевых сегментах.

Литература

1. Уэбстер Ф. Основы промышленного маркетинга. М.: Изд. дом Гребенникова, 2005. 416 с.
2. Ламбен Ж.-Ж. Менеджмент, ориентированный на рынок. Стратегический и операционный маркетинг / Пер. с англ. СПб.: Питер, 2007. 800 с.
3. Метрики управления портфелем программных продуктов / Б.Ходжес [и др.] // Открытые системы. 2007. № 3. [Электронный ресурс]: http://www.osp.ru/os/2007/03/4177900/ (дата обращения 09.02.2015).
4. Дубова Н. Управление с портфелем // Открытые системы. 2008. № 3. [Электронный ресурс]: http://www.osp.ru/os/2008/03/5015107/ (дата обращения 09.02.2015).
5. Оганесян А. Опыт управления портфелем ИТ-проектов // Открытые системы. 2008. № 3. [Электронный ресурс]: http://www.osp.ru/os/2008/03/5015181/ (дата обращения 09.02.2015).
6. Авдошин С.М., Лифшиц А.А. Формирование портфеля проектов на основе нечеткой модели многокритериальной оптимизации // Бизнес-информатика. 2014. № 1 (27). С. 14–22.
7. Недовесов М.В., Руденко З.Г. Формирование оптимального портфеля взаимозависимых проектов и его оптимизация по времени // Проблемы управления. 2012. № 4. С. 26–31.
8. Kang K., Lee J., Donohoe P. Feature-oriented product line engineering // IEEE Software. July/August 2002. P. 58–65.
9. Кознов Д.В., Новицкий И.А., Смирнов М.Н. Инструменты для управления вариативностью: готовность к промышленному применению // Труды СПИИРАН. 2013. № 3 (26). С. 297–331.
10. Гусс С.В. Разработка семейства программных систем в специфической предметной области // Математические структуры и моделирование. 2011. № 22. С. 55–68.
11. Michalek J.J., Ebbes P., at all. Enhancing marketing with engineering: Optimal product line design for heterogeneous markets // International Journal of Research in Marketing. 2011. No. 28. P. 1–12.
12. Костерин А.Г. Практика сегментирования рынка. СПб.: Питер, 2002. 288 с.
13. Ехлаков Ю.П., Бараксанов Д.Н., Мамонова Н.В. Функциональная и математические модели сегментирования рынка программных продуктов // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2012. № 2. С. 155–160.
14. Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. М.: Советское радио, 1975. 192 с.
15. Вуд М.Б. Маркетинговый план. Практическое руководство по разработке. М.: Вильямс, 2007. 352 с.
16. Анализ ассортимента с помощью матрицы McKinsey – General Electric [Электронный ресурс]: http://powerbranding.ru/biznes-analiz/matrica-mckinsey/ (дата обращения 20.06.2014).
17. Что такое привлекательность рынка и как ее оценить? [Электронный ресурс]: http://powerbranding.ru/rynok/ocenka-privlekatelnosti/ (дата обращения 20.06.2014).
18. Фатрелл Р.Т., Шафер Д.Ф., Шафер Л.И. Управление программными проектами. Достижение оптимального качества при минимуме затрат. М.: Вильямс, 2004. 1136 с.
19. Российское образование: справочник. [Электронный ресурс]: http://www.edu.ru/ (дата обращения 09.02.2015).

ANNOTATED SUFFIX TREE AS A WAY OF TEXT REPRESENTATION FOR INFORMATION RETRIEVAL IN TEXT COLLECTIONS

Dmitry S. FROLOV

Post-graduate student, Department of Data Analysis and Artificial Intelligence,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: dfrolov@hse.ru, dmitsf@gmail.com

A method for information retrieval based on annotated suffix trees (AST) is presented. The method is based on a string-to-document relevance score calculated using AST as well as fragment reverse indexing for improving performance. We developed a search engine based on the method. This engine is compared with some other popular text aggregating techniques: probabilistic latent semantic indexing (PLSA) and latent Dirichlet allocation (LDA).

We used real data for computation experiments: an online store's xml-catalogs and collections of web pages (both in Russian) and a real user's queries from the Yandex. Wordstat service. As quality metrics, we used point quality estimations and graphical representations. Our AST-based method generally leads to results that are similar to those obtained by the other methods. However, in the case of inaccurate queries, AST-based results are superior. The speed of the AST-based method is slightly worse than the speed of the PLSA/LDA-based methods. Due to the observed correlation between the average query performing time and the string lengths at the AST construction phase, one can improve the performance of the algorithm by dividing the texts into smaller fragments at the preprocessing stage. However, the quality of search may suffer if the fragments are too short. Therefore, the applicability of annotated suffix tree techniques for text retrieval problems is demonstrated. Moreover, the AST-based method has significant advantages in the case of fuzzy search.

Introduction

One of the key areas in data analysis is the processing of document collections: document rubrication, text similarity scoring, document search by given keywords and so on [4, 5, 9, 11]. The last problem forms a particular part of computer science, so-called information retrieval. Information retrieval problems are widely covered in books and research papers. A major reference is 'Introduction to Information Retrieval' by C.Manning et al. [11], which can be used as a textbook for information retrieval studies as well. In papers [4, 9], belonging to a rather popular scientific genre, one may find many methods and techniques which are applicable for design and implementation of search engines. In paper [4] information retrieval tools are considered, including retrieval techniques for specific documents like Internet forums pages at which most information is produced by users.

In principle, the problem of document retrieval for a given query can be considered as a task of full-text search or even as a basic string matching problem. There have been a number of methods developed for this problem including that emphasized in paper [15]. Applicable techniques aim at a reasonable balance between quality of obtained results and retrieval time.

A classic approach in information retrieval is indexing of document collections [5, 11]. However, there is a need to improve performance and accuracy, which leads to developing alternative approaches. One of these alternative methods is aggregate representation both for individual texts [3, 13, 14] and for collections as well [2, 7]. The Annotated suffix tree method (AST) was proposed in papers by B.G.Mirkin and E.L.Chernyak [3, 14]. The authors use AST for determining the so-called string-to-document relevance score and automated construction of taxonomy systems [3]. Regarding other approaches to aggregate document collection representations, let us indicate, first of all, probabilistic latent semantic indexing (PLSI) [7] and latent Dirichlet allocation (LDA) [2]. A recent survey of these and similar techniques can be found in A.Korshunov and A.Gomzin's paper [8]. Further modifications of classic methods in probabilistic topic modelling can be found in K.V.Vorontsov's papers, for example, in [16].

Note that both the probabilistic latent semantic indexing and latent Dirichlet allocation use feature-based text representation, whereas the annotated suffix tree approach uses fragment representation only [13]. This feature is very important for fuzzy text searching, for example, if the search query or documents in a collection contain mistakes. Feature-based methods require special adaptation for such situations [11]. The purpose of this work is to propose a technique for applying AST in information retrieval of texts or queries containing mistakes and to analyze its efficiency against the methods using probabilistic topic modelling.

Currently there are several implementations of the AST method. Using annotated suffix arrays (ASA) and tree constructing algorithms for substring search proposed in [6] significantly reduces the time complexity of text aggregation and provides an appreciable saving of memory resources.

1. Methods

In this work we compare aggregate text representations of whole document collections (PLSI, LDA) and individual texts (AST) as ways to solve the information retrieval problem for a given query.

In PLSA and LDA information retrieval, the problem is reduced to a problem of detecting documents which are similar to the given one (search query). Therefore the search query is transformed into a feature vector form used in these models [11]. After that, one has to carry out procedures of calculating the similarity score between the query and representations of all documents in a collection [17], sorting by obtained values and choosing documents gaining the greatest scores. Of course, different parameters of the models (in particular, the number of detected themes and algorithm parameters) may lead to different results of these procedures. Note that there are many implementations of PLSA and LDA models; the main differences between these implementations are development tools. For our experiments, we used implementations from the free distributed framework Gensim (*https://radimrehurek.com/gensim*) for Python, with significant modifications and special adjustments for our aims. In particular, the feature processing module was appended, and we also changed document storage formats for more convenient work with the databases we used.

Let us describe a method we developed for full-text search based on AST representations of the documents. Here we give a classical AST constructing algorithm for a given document [3]. AST is a weighted root tree, which is used as a data structure for storing and processing text, which stores fragments and corresponding frequencies. One of the tree nodes is a root of the tree; an empty tree contains the root only. Other nodes contain text symbols and corresponding frequencies (so-called annotations).

To use AST, we have to split document into strings — symbol sequences. As a rule, one string is formed by 2-4 sequential words from a text. A k-suffix of a string $x = x_1 x_2 ... x_N$ of length N is a substring $x_k = x_{N-k+1} x_{N-k+2} ... x_N$.

For example, 3-suffix of a string *INFORNATION* is a substring *ION*. Note that an N-suffix of any string is a whole string. An algorithm for constructing AST for a given string x is described below:

1. Initialize an empty AST T;

2. Find all suffixes of a given string $\{x_k : k = 1, 2, ... N\}$;

3. For each suffix x_k find maximal overlap (path from the root) in T: x_k^{max}. For all nodes from x_k^{max} increase annotations by 1. If the length of x_k^{max} is less than k, one creates new nodes from the remaining part of this suffix; annotations of all the new nodes are equal to 1.

An example of AST for a string *ABCBA* is shown in *Fig. 1.*

To construct AST for 2 or more strings, one has to perform sequential addition of these strings into a common tree. Hence, if we represent a document as a set of strings, we can construct an AST for the document.

The AST constructing algorithm, which is given above, is quite costly both for the time and space complexity. Currently there are methods based on classical string algorithms, for example, the Ukkonen algorithm
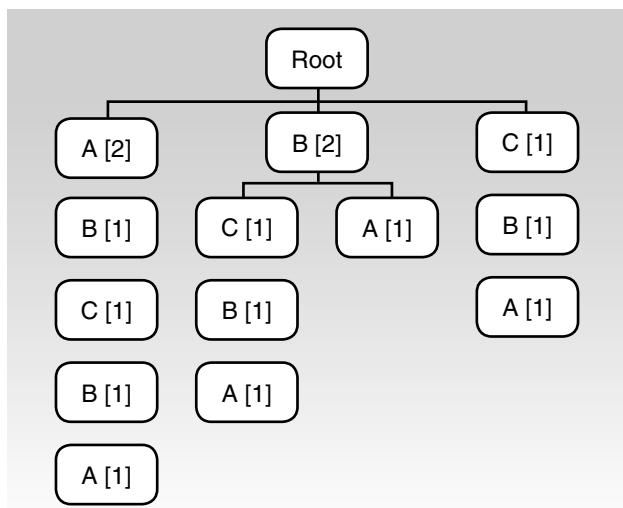
Fig 1. An example of AST for a string ABCBA

for substring matching [6]. This makes it possible to reduce time complexity for AST construction. Another way to improve the algorithm is to reduce memory consumption. This goal can be achieved using special data structures for AST storage, for example, special arrays.

The AST constructed for the document allows us to solve a problem of similarity scoring between a given string and the document. This problem is called string-to-document relevance scoring. Let us describe a way to solve this problem proposed in [3].

Consider the conditional probability of node u in given AST T with the root R. Denote annotation of node u as $f(u)$. The conditional probability of node u is:

$$p(u) = \frac{f(u)}{\sum\limits_{v \in T: ancestor(v) = ancestor(u)} f(v)}, \qquad (1)$$

where $ancestor(u)$ — is an ancestor of node u in the AST T.

For each suffix x_k of the string x, the relevance score is defined as the sum of conditional probabilities of all nodes which belong to maximal overlap x_k^{max}:

$$s(x_k, T) = \frac{1}{k_{max}} \sum\limits_{i=0}^{k_{max}} p(x_i^k). \qquad (2)$$

Finally, the relevance score of the string x for AST T is expressed by the following formula:

$$S(x, T) = \frac{1}{N} \sum\limits_{k=0}^{N} s(x_k, T). \qquad (3)$$

For example, the relevance score of a substring *BAC* and AST constructed for a string *ABCBA* is 0.35.

Hence, one can formulate the following strategy for retrieval documents for a given query. If we have an AST constructed for each document in a collection, then one can calculate string-to-document relevance scores, sort them and obtain the most relevant documents. Obviously, iteration of the procedure for string-to-document relevance scoring makes the speed of this method very slow. We propose to compute relevance scores not for a whole collection but for documents which have a high probability to gain a high score. To detect these documents, we use an approach based on an inverted index [5, 11] of ancillary features (fragments). We use a special data structure, which is exactly a hash-table as the following:

$$\begin{cases} f_1 : \left[n_{11}, ..., n_{1m_1} \right] \\ f_2 : \left[n_{21}, ..., n_{2m_2} \right] \\ ... \\ f_K : \left[n_{K1}, ..., n_{Km_K} \right] \end{cases}, \qquad (4)$$

where f_i — document fragments;

n_{ij} — indexes of corresponding documents (or links to documents) which contain these fragments;

K — count of fragments.

We can consider any document feature as a fragment. In our implementation, we use 3- and 4-grams (sequential symbols of a text). One can use individual words or bigrams; this approach gives us a classical inverted index. If a collection is very large, one can consider not all fragments but fragments having a frequency which is more than the given one. Thus, query processing before string-to-document relevance scoring includes the selection of «candidate documents».

The final search algorithm consists of the following steps:

1. Split a query into substrings and select corresponding candidate documents from an inverted index;

2. Calculate the string-to-document relevance score for these candidate documents using AST;

3. Sort the documents by relevance score.

These steps are the simplest indexing algorithm before using AST. One can improve the hash-table using fragments with the numbers of their occurrences in documents, of some transformations (for example, tf-idf [12]). In principle, it is possible to use ranking functions (BM25 and others) with a hash-table and following comparison with the results obtained with AST. A particular issue to investigate is a situation if there are no ancillary features in a search query. In our algorithm, we include a whole collection in a list of candidate documents.

2. Experiments and results

Let us describe an experimental approval of the proposed method. In our comparison, we have considered the AST-based method and methods based on PLSI and LDA (for the last one we implemented a modification used bigrams [2]). We used a laptop with a 2.0 GHz dual-core processor and 4 GB RAM, running under the Ubuntu 12.04 operating system for our aims. Test document collections were stored in the document-oriented database MongoDB (*http://mongodb.org*). The full-text search engine of this database was also included in our comparison. We used 2 document collections for experimental approval.

The first (Collection #1) was produced from the xml-catalogue of fantasy books in Russian in the Ozon.ru on-line superstore (*http://www.ozon.ru*). The catalogue contains about 90,000 documents, formatted to contain the title, author, description and a set of parameters (such as current price, discount, number of pages, publisher, etc.). To use these documents for our purposes, we remove all except the title, author and description and transform documents into text form.

We used real user queries obtained from the Yandex. Wordstat online service (*http://wordstat.yandex.ru*). This service is used for efficient website promotion in the Yandex search engine and SEO-optimization. It allows us to get frequency statistics for a given query (weekly, monthly etc.) and to obtain a wide list of associated queries (which are similar by form or by search results). We used the last feature. To obtain a set of queries from this system, we make requests for some chosen test query q and obtain a list of real user queries $S(q) = \{q_1, ..., q_N\}$. We formed 2 sets of test queries. The first one contained titles of subcategories in the original xml-catalog; the second one consisted of words and phrases from documents. From these sets of queries was formed 3 groups of real users' queries. Group #1 consisted of queries from lists $S(q)$ for the first set. We called this group «Titles of subcategory». Groups #2 and #3 - «Transparent queries» and «Inaccurate queries» consisting of user queries obtained from lists for the second set. But for Group #3, we selected inaccurate and damaged queries in $S(q)$ lists. Altogether we considered 90 queries. Note that we used special scripts in Python to automate processing of the results of our experiments.

To measure quality metrics of the methods under consideration, precision at level of N documents R_N and recall R were calculated. These metrics are calculated for each query using cardinality of the relevant document set and set of documents retrieved by the system. Note that one has to choose N documents (which has the largest weight assigned by the system) as a second set to calculate N document level precision. The standard formulas are as follows:

$$R = \frac{\left| X_{relevant} \cap X_{obtained} \right|}{\left| X_{relevant} \right|}, \qquad (5)$$

$$P_N = \frac{\left| X_{relevant} \cap X_{obtained}^{N} \right|}{\left| X_{obtained}^{N} \right|}, \qquad (6)$$

where $X_{relevant}$ — a set of documents which are relevant for a given query;

$X_{obtained}$ — set of documents retrieved by the system;

$X_{obtained}^{N}$ — set of N documents which gained the largest relevance score in the search system for a given query.

Average precision of the methods is shown in *Table 1* and *2* for the level of documents 5 and 10 respectively.

Table 1.

5-document level precision

Number of group of queries	AST	PLSI	LDA	LDA with bigrams	MongoDB fulltext search
1	**0.87**	0.71	0.85	**0.87**	0.55
2	0.84	0.64	0.82	**0.86**	0.57
3	**0.78**	0.43	0.45	0.59	0.28
Average	**0.83**	0.59	0.71	0.77	0.47

Table 2.

10-document level precision

Number of group of queries	AST	PLSI	LDA	LDA with bigrams	MongoDB fulltext search
1	0.85	0.70	0.85	**0.86**	0.5
2	0.84	0.68	0.81	**0.86**	0.5
3	**0.79**	0.41	0.43	0.55	0.2
Average	**0.82**	0.56	0.70	0.76	0.4

Note that precision levels of the AST-based method for the 3rd group of queries («Inaccurate queries») are almost the same as for the 1st and 2nd groups. This should be attributed to the fact that AST uses text fragments, not features.

To show the level of recall and see the interplay between the precision and recall, let us use 11-point precision-recall TREC curves [1]. Such a curve is one of the graphical representations of the quality level of a search engine. It shows the precision level for a given recall level at the same search engine. Usually one uses 11 levels of
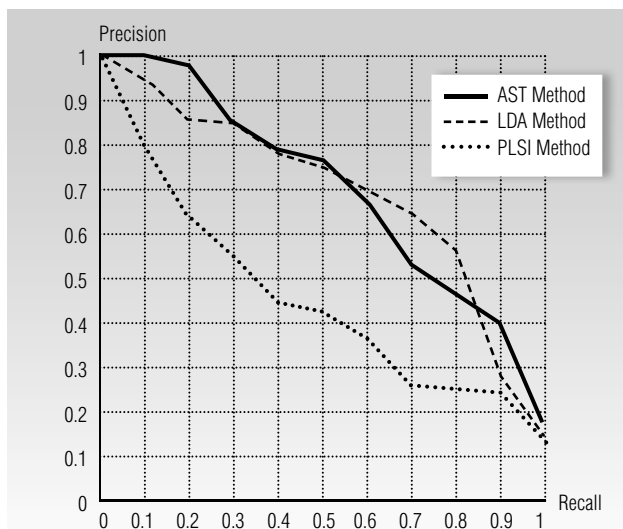
Fig. 2. Precision-recall curves for different methods, group of queries #1
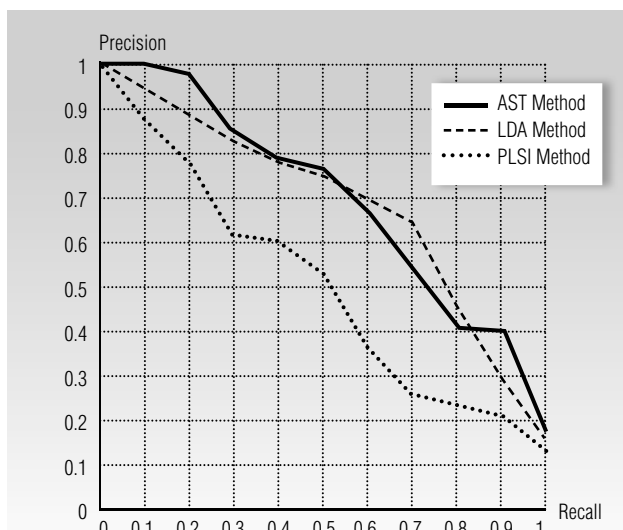
Fig. 3. Precision-recall curves for different methods, group of queries #2
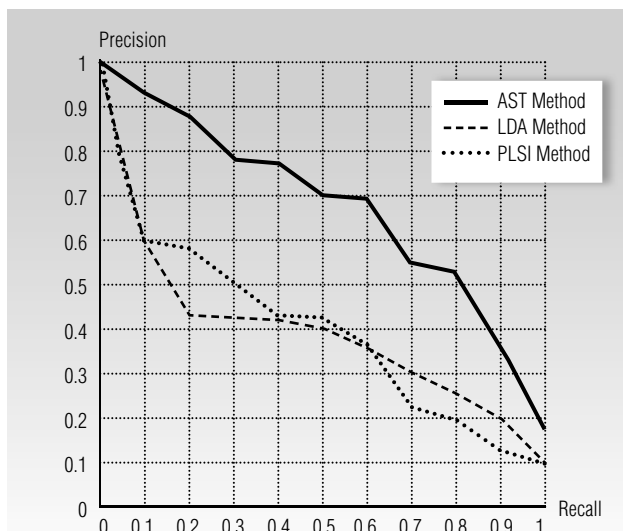
Fig. 4. Precision-recall curves for different methods, group of queries #3

recall (from 0.0 to 1.0 with a step equal to 0.1). It gives more complete information than one point characteristic. Les us describe a procedure of constructing the precision-recall curve.

1. Fix 11 recall values from 0.0 to 1,0 with the step 0.1;
2. Choose and fix the set of queries;
3. For each query perform the precision interpolation procedure for given recall values R_i, $i = 1, ..., 11$:
   - carry out the search procedure and obtain a list of results sorted by descending relevance score;
   - in this list note the number of the document for which the value of recall is equal to $R_i - K$, and calculate the corresponding precision value (for level of K documents) - P_K;
4. Find the averages for obtained values;
5. Get the pairs of values (precision, recall), draw a curve.

Precision-recall curves for different search methods obtained for different group of queries (1, 2, 3) are shown by *Fig. 2-4* below.

For the first two groups, the AST-based method leads to results that are similar to those obtained by the other methods. However, for the 3rd group of queries («Inaccurate queries») the AST method results are superior.

Average query processing times are shown in *Table 3*. So, the AST-based method yields to PLSA and LDA but nevertheless it is significantly faster than MongoDB full-text search. Note that the performance of the AST method is a really difficult problem.

Table 3.

Average time of query processing for different methods (s)

AST	PLSI	LDA	LDA with bigrams	MongoDB fulltext search
0.43	**0.23**	0.25	0.29	8.55

The second collection of documents was obtained from a set of Habrahabr.ru (*http://habrahabr.ru*) web pages. We downloaded several thousand web pages from this site and chose the 2,000 largest pages from this set. After that we removed the HTML markdowns from the web pages and saved the results as raw texts. The average length of these texts was significantly greater than of those used in the first experiment.

We use this collection of documents in order to test the performance of the search engines under consideration. The queries here were chosen as slightly modified strings of the documents. The AST-based retrieval method appears to be slower than those based on PLSI and LDA, but it is significantly faster than the MongoDB embedded full-text search engine.

Table 4.

Average time of query processing for different methods (s)

AST	PLSI	LDA	LDA with bigrams	MongoDB fulltext search
0.15	**0.04**	0.05	0.08	3.09

The correlation between the time for query processing and length of strings used for AST constructing also was investigated. As a rule, one uses 2-3 sequential words as a string for tree constructing. We implement algorithm modifications using 1 word and 5-6 words as a string. The experiment results are presented in *Table 5*. From the results one can conclude that string «extending» increases the query processing time.

Table 5.

Influence of string length used in AST and query processing time (s)

AST with 1 word per string	AST with 2-3 words per string	AST with 5-6 words per string
0.12	0.15	0.21

One can suppose we can improve the performance of the algorithm by using the minimal possible length of AST strings. However, in such a case we can lose helpful conjunctions between words and impair the quality metrics of the algorithms. We are going to investigate this in further work.

Conclusion

We described and verified the method of applying AST technique for information retrieval in document collections. Experiments using real data show us that the AST-based method does not yield to classical PLSI and LDA methods, and, furthermore, has some advantages, for example, in the case of inaccurate queries. We can conclude that it is necessary to make further investigations of the applicability of AST techniques in information retrieval. Our future plans include developing effective text retrieval methods based on the AST techniques. This includes the following subtasks:

1. Improvement of the AST-based method for document retrieval with the help of both index-based principles and approaches using no indexing;

2. Developing a method for document retrieval using aggregate representation of a whole collection or representation of document groups;

3. Adaptation of the methods for distributed document storing and computation;

4. Adaptation of the methods to cases of dynamically changing document collections;

5. Conducting experimental computations for comparison of the developed methods.

I wish to thank Prof. B. Mirkin for his advice and valuable comments. ∎

References

1. Ageev M., Kuralenok I., Nekresyanov I. (2004) Prilozhenie A. Oficial'nye metriki ROMIP'2004 [ROMIP'2004 official metrics]. Proceedings of the *ROMIP 2004 Seminar, October 1, 2004, Saint Petersburg, Russia*. Saint Petersburg: SPbSU Institute of Chemistry, pp. 142–150 (in Russian).

2. Blei D.M., Ng A.Y., Jordan M.I. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research*, no. 3, pp. 993–1022.

3. Chernjak E.L., Mirkin B.G. (2013) Ispol'zovanie resursov Interneta dlja postroenija taksonomii [Using the Internet for Taxonomy Constructing]. Proceedings of the *AIST 2013 Conference, April 4-6, 2013, Ekaterinburg, Russia*. National Open University INTUIT, pp. 36–48 (in Russian).

4. Croft W.B., Metzler D., Strohman T. (2010) *Search engines: Information retrieval in practice*, Boston: Addison-Wesley.

5. Cutting D., Pedersen J. (1989) Optimization for dynamic inverted index maintenance. Proceedings of the *13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGR 1989), June 25-28, 1989, Cambridge, MA, USA*. NY: ACM Press, pp. 405–411.

6. Dubov M., Chernjak E. (2013) Annotirovannye suffiksnye derev'ja: Osobennosti realizacii [Annotated suffix trees: Implementation features]. Proceedings of the A*IST 2013 Conference, April 4-6, 2013, Ekaterinburg, Russia*. National Open University INTUIT, pp. 49–57 (in Russian).

7. Hofmann T. (1999) Probabilistic latent semantic indexing. Proceedings of the *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1999). August 15-19, 1999, Berkeley, CA, USA*. NY: ACM Press, pp. 50–57.

8. Korshunov A., Gomzin A. (2012) Tematicheskoe modelirovanie tekstov na estestvennom jazyke [Topic modeling for texts in natural language]. *Proceedings of the Institute for System Programming of the RAS*, no. 23, pp. 215–238 (in Russian).

9. Langville A. N., Meyer C. D. (2011) *Google's PageRank and beyond: The science of search engine rankings*, Princeton: Princeton University Press.

10. Malkov Yu., Ponomarenko A., Logvinov A., Krylov V. (2014) Approximate nearest neighbor algorithm based on navigable small world graphs. *Information Systems*, no. 45, pp. 61−68.

11. Manning K., Raghavan P., Shjutce H. (2011) *Vvedenie v informacionnyj poisk* [Introduction to information retrieval]. Moscow: Williams (in Russian).

12. Mirkin B. (2011) *Core concepts in data analysis: Summarization, correlation and visualization*, London: Springer.

13. Mirkin B.G. (2011) *Metody klaster-analiza dlya podderzhki prinyatiya reshenii: obzor* [Cluster analysis for decision making: Review]. Working paper WP7/2011/03. Moscow: HSE (in Russian).

14. Mirkin B.G., Chernjak E.L., Chugunova O.N. (2012) Metod annotirovannogo suffiksnogo dereva dlja ocenki stepeni vhozhdenija strok v tekstovye dokumenty [Annotated suffix tree as a way of string-to-document score evaluating]. *Business Informatics*, no. 3 (21), pp. 31−41 (in Russian).

15. Natarajan J., Shaw S., Bruchez R., Coles M. (2012) *Pro T-SQL 2012 Programmer's Guide*, NY: Apress.

16. Vorontsov K.V., Potapenko A.A. (2011) Modifikacii EM-algoritma dlja verojatnostnogo tematicheskogo modelirovanija [EM algorithm modifications for probabilistic topic modeling]. *Journal of Mathematical Physics*, no. 52, pp. 1−21 (in Russian).

17. Wei X., Croft W.B. (2006) LDA-based document models for ad-hoc retrieval. Proceedings of the *29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGR 2006), August 6-11, 2006, Seattle, WA*, USA. NY: ACM Press, pp. 178−185.

ПРИМЕНЕНИЕ МЕТОДА АННОТИРОВАННОГО СУФФИКСНОГО ДЕРЕВА В ЗАДАЧАХ ПОИСКА В КОЛЛЕКЦИЯХ ТЕКСТОВЫХ ДОКУМЕНТОВ

Д.С. ФРОЛОВ
*аспирант, департамент анализа данных и искусственного интеллекта,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: dfrolov@hse.ru, dmitsf@gmail.com*

В работе представлен метод информационного поиска в коллекциях текстовых документов, основанный на аннотированных суффиксных деревьях (АСД). В методе используется определение степени вхождения строки в АСД, полученные для документов, а также обратный индекс, построенный по фрагментам документов (с целью улучшения производительности). На основе представленного метода реализована поисковая система и произведено ее сравнение с алгоритмами поиска, использующими другие способы агрегированного представления текстов (всей коллекции целиком) — вероятностным латентно-семантическим индексированием (PLSI) и скрытым размещением Дирихле (LDA).

Для проведения вычислительных экспериментов использованы реальные данные: коллекция xml-каталогов онлайн-магазина и коллекция веб-страниц (обе – на русском языке), а также пользовательские поисковые запросы, полученные с помощью сервиса Yandex.Wordstat. Исследованы качественные метрики рассматриваемых систем: получены точечные оценки и графические характеристики. Метод поиска, основанный на АСД, в целом показывает результаты, сравнимые с другими алгоритмами, однако, на неточных запросах существенно превосходит их. Была исследована производительность сравниваемых поисковых систем, в результате отмечено, что метод на основе АСД несколько уступает другим по скорости поиска. Также изучена зависимость между временем выполнения запроса и длиной строк текста, используемых для построения АСД: для улучшения производительности необходимо выбирать минимально возможную длину строк, принимая во внимание тот факт, что слишком короткие строки могут ухудшить качественные характеристики метода. Отдельно отмечен факт применимости метода на основе АСД к задачам нечеткого поиска, что должно стать предметом будущих исследований.

Литература

1. Агеев М., Кураленок И., Некрестьянов И. Приложение А. Официальные метрики РОМИП'2004 // Труды второго российского семинара по оценке методов информационного поиска, 1 октября 2004 г., г. Санкт-Петербург; под ред. И.С.Некрестьянова. СПб.: Институт Химии СПбГУ, 2004. С. 142—150.

2. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet allocation // Journal of Machine Learning Research. 2003. No. 3. P. 993—1022.

3. Черняк Е.Л., Миркин Б.Г. Использование ресурсов Интернета для построения таксономии // Доклады всероссийской научной конференции АИСТ 2013, 4-6 апреля 2013 г., г. Екатеринбург. Национальный открытый университет ИНТУИТ, 2013. С. 36—48.

4. Croft W.B., Metzler D., Strohman T. Search engines: Information retrieval in practice. Boston: Addison-Wesley, 2010. 520 p.

5. Cutting D., Pedersen J. Optimization for dynamic inverted index maintenance // Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGR 1989), June 25-28, 1989, Cambridge, MA, USA. NY: ACM Press, 1989. P. 405—411.

6. Дубов М., Черняк Е. Аннотированные суффиксные деревья: особенности реализации // Доклады всероссийской научной конференции АИСТ 2013, 4-6 апреля 2013 г., г. Екатеринбург. Национальный открытый университет ИНТУИТ, 2013. С. 49—57.

7. Hofmann T. Probabilistic latent semantic indexing // Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1999). August 15-19, 1999, Berkeley, CA, USA. NY: ACM Press, 1999. P. 50—57.

8. Коршунов А., Гомзин А. Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. 2012. № 23. С. 215—238.

9. Langville A.N., Meyer C.D. Google's PageRank and beyond: The science of search engine rankings. Princeton: Princeton University Press, 2011. 240 p.

10. Approximate nearest neighbor algorithm based on navigable small world graphs / Malkov Yu. [et al.] // Information Systems. 2014. No. 45. P. 61—68.

11. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск. М.: Вильямс, 2011. 528 с.

12. Mirkin B. Core concepts in data analysis: Summarization, correlation and visualization. London: Springer, 2011. 390 p.

13. Миркин Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор. Препринт WP7/2011/03. М.: НИУ ВШЭ, 2011. 39 с.

14. Миркин Б.Г., Черняк Е.Л., Чугунова О.Н. Метод аннотированного суффиксного дерева для оценки степени вхождения строк в текстовые документы // Бизнес-информатика. 2012. № 3 (21). С. 31—41.

15. Pro T-SQL 2012 Programmer's Guide / M.Coles [et al.]. NY: Apress, 2012. 679 p.

16. Воронцов К.В., Потапенко А.А. Модификации EM-алгоритма для вероятностного тематического моделирования // Journal of Mathematical Physics. 2011. № 52. С. 1—21.

17. Wei X., Croft W.B. LDA-based document models for ad-hoc retrieval // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGR 2006), August 6-11, 2006, Seattle, WA, USA. NY: ACM Press, 2006. P. 178—185.

SECURING INFORMATION RESOURCES USING WEB APPLICATION FIREWALLS

Petr A. BARANOV
Associate Professor, Department of Innovations and Business in IT,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: pbaranov@hse.ru

Eldar R. BEYBUTOV
MSc Program Student, Faculty of Business and Management,
National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation
E-mail: eldar.beybutov@gmail.com

This paper provides an overview of core technologies implemented by comparably new products on the information security market – web application firewalls. Web applications are a very widely-used and convenient way of presenting remote users with access to corporate information resources. They can, however, become single point of failure rendering all the information infrastructure inaccessible to legitimate clients. To prevent malicious access attempts to endpoint information resources and, intermediately, to web servers, a new class of information security solutions has been created.

Web application firewalls function at the highest, seventh layer of the ISO/OSI model and serve as a controlling tunnel for all the traffic heading to and from a company's web application server(s). To ensure decent levels of traffic monitoring and intrusion prevention, web application firewalls are equipped with various mechanisms of data exchange session «normality» control. These mechanisms include protocol check routines, machine learning techniques, traffic signature analysis and more dedicated means, such as denial of service, XSS injection and CRRF attack prevention. The ability to research and add user rules to be processed along with vendor-provided ones is important, since every company has its own security policy and, therefore, the web application firewall should provide security engineers with ways to tweak its rules to reflect the security policy more precisely.

This research is based on broad practical experience of integrating web application firewalls into the security landscape of various organizations, their administration and customization. We illustrate our research into available filtering mechanisms and their implementations with exemplary product features by market leaders.

Key words: information security, web application firewall, application server protection.

Introduction

Nowadays, many companies and businesses have an information security policy which assumes remote access to their information resources (calculation powers, cloud services, data storage). By saying «remote access», we mean access over the Internet. This could be made through letting all Internet users gain certain types of access to resources or through letting only identified corporate users remotely use the company's resources. Direct access to information resources is both inconvenient for users and comparably insecure because of the lack of a single access point

which hinders security policy implementation. The solution is well-known and it is called Application Servers, or application-layer intermediate nodes. External users gain access to these nodes using the regular web browser, interact with unified interface and put queries to it. These queries are afterwards translated by the Application Server into more specific queries to internal information resources and, after getting a response from these resources, the Application Server transforms them into an easy-to-understand view and shows it to the external user in his/her web browser. The scheme is transparent and, once all components are installed and set up, has predictable and controllable technical support expenses. Information security breaches often lead to increased technical support expenses. In these terms, to gain control over technical support funds the company management must be sure that the security level of the single access point – the web application – is high enough to prevent malicious attempts at access and use of company data from getting through the Application Server to the data hosting infrastructure. There are national [1, 2], industry branch [3, 4] and corporate [5, 6] standards of writing secure web applications. Application development in accordance with these standards is a labor-intensive, expensive and hard-to-scale procedure. It does not guarantee safety of the result if software developed by a third party is applied. Information security officers need a versatile and configurable tool to control traffic flowing through the web application server and it must be able to prevent data endpoints and the application server itself from receiving and processing maliciously crafted traffic and queries.

Web application firewalls present a solution for the problem described. There are numerous vendors offering a variety of products who claim their products have all the mechanisms needed to provide security on top of standard web application rules. Speaking of Web application firewalls (WAF), people often become confused because of the different associations they have in mind on what features such a specific tool should contain. Even key functions of WAF are sometimes misunderstood. In this research, we would like to introduce a WAF's typical functionality and the defense mechanisms that are essential for WAF in the modern state of the industry. The research is based on practical experience of integrating WAFs of different vendors into existing company information infrastructures.

First of all, we will define a list of the defense mechanisms for a WAF which must be present. We will describe every mechanism, its features and how it works. To illustrate the way some mechanisms function, we will make comparisons of their implementation for different WAF vendors. As examples, we will mainly focus on solutions from leaders in the area their solutions prove to be quite representative (for more about WAF market leaders see [7]). The question of «which WAF is the best» has no answer, and we have chosen these vendors just to highlight variations of implementation of a single product ideology in different practices.

Let's list the most necessary defense mechanisms for any WAF:

✦ protocol check;

✦ signature analysis;

✦ machine learning of access identifier formats;

✦ injection and XSS protection (mostly proprietary);

✦ user-defined rules of illegitimate queries detection;

✦ «denial of service» attack prevention;

✦ integration of the information security landscape.

We will now go through all of these mechanisms and explain their capabilities.

1. Protocol check

Protocol check is a passive protection mechanism against potential threats exploiting non-typical use of HTTP protocol features. Firstly, it involves HTTP header check for compliance with RFC. But RFC doesn't have all the rules and restrictions to ensure security of the traffic and vendors invent their own restrictions that, of course, do not interfere with the work of legitimate users, once implemented.

Protocol check is one of the primary mechanisms. Its main role is to leave the intruder as few opportunities as possible to exploit possible internal vulnerabilities. The HTTP transactions are limited through the following checks:

◇ RFC requirements;

◇ header and parameter's length and number;

◇ time limits;

◇ JSON and XML entity checks;

◇ illegal values detection.

2. Signature analysis

Signature analysis is one of the eldest technologies of ensuring the security of applications. It is still widely used and its effectiveness is proven through the number of devices and solutions based on it (e.g. antivirus software, classic firewalls, spam-filters).

Modern trends in the evolution of information security threats show that nowadays the majority of intruders do not develop their own maliciously-oriented software. During a typical attack, a malefactor applies ready-to-use hacking means previously created by a third party (for examples see [8]. Moreover, the usage intensity for these means is so high that public web applications suffer automated attacks nearly all the time. Theoretically, mechanisms involving a machine learning process that creates a normal behavior model can render signature analysis useless. Based on this fact, some WAF vendors do not rely fully on the signature analysis mechanism in their products and do not invest much in signature update procedures (for example instead of analyzing User-Agent and other signature-based methods, Wallarm WAF implements behavioral fingerprinting schemes to determine the tools used in the attack [9]).

But practice shows that in some cases this protection mechanism is irreplaceable. For example, during the machine learning period signatures prove to be very useful due to the fact that they ensure a «clean» data environment for anomaly-detection software. They also ensure the overall security level for the leaning period. Because of this, a decent WAF solution must have a wide and relevant signature database applicable for all types of web applications.

3. Machine learning

Machine learning of access identifier formats is one of the key features for products of WAF-class. The main concept here is creation of a normal behavior model based on URL, parameters and cookies. Once a model is created and tested, the comparison of live traffic against it could prevent both known and unknown vulnerabilities from being exploited. Let's try to estimate the efficacy of machine learning.

It must be said, that the technology's efficiency is hard to estimate by means of a mathematical statistical algorithm because the algorithm is usually proprietary and is not disclosed by solution producers. Live traffic imitation poses certain difficulties as well.

Nonetheless, we are able to compare the following features of the algorithms:

◆ flexibility of learning parameters;

◆ resulting data optimization.

In a solution from F5, for example, for every web application being protected a profile is created. This profile specifies the beginning and ending of learning thresholds. There is a possibility to list an interval of trusted IP addresses. It is very important, since queries from such IP addresses would represent an invaluable contribution to a concept of the normal traffic model. As for machine learning improvement, the replies from web applications are also analyzed: the parameters used in reply forms are more trusted for WAF than those created by a remote client.

This functionality is more «boxed» when speaking of Imperva's solution. The only parameterized value is the query learning time limit. By default, it is set to 240 hours for all web applications. This means that after 240 hours of learning process of an object (application), the object is treated as «learned.» All the queries from now on will be compared in relation to the normal model.

The resulting data optimization is a vital process. The need to perform it is raised when the normal model is not formed correctly or the web application being protected has been modified by its developers. Most WAF products support this feature to be performed by hand (it is called «manual object adjustment»). Some of them provide users with automatic object adjustment.

Usually automatic measures here are implemented through a mechanism of tracking changes. It monitors the number of false positive events generated by the model. For example, for the WAF by F5, if 5 different users in a 5 minute time interval performed the same violation, the object is switched to re-learning state. Another example used in the Imperva solution tracks the number of typical model violations per 12 hours. If the number is higher than 50 for every hour throughout a 12-hour interval, the object is rendered «unlearned».

4. Injection
nd XSS protection

Positioned between the web application server and the outer network environment, WAF as a security tool has an opportunity to «comprehend» the traffic going through it, analyze it and check it for compliance with security rules. Injection attacks take place in cases where the web application sends unchecked (or not sufficiently checked) data taken from client's query to a neighboring system's command interpreter. Neighboring systems here could be databases, the operating system, LDAP-server, XPath interpreter and many others. This query transmission allows malefactors to manipulate adjacent functional systems.

Injection prevention is achieved through application of the following mechanisms:

✧ tokenization. Using a finite automaton, the query is parsed and the target system's tokens are detected. When certain (previously defined in WAF parameters) tokens are found, the query is treated as potentially dangerous;

✧ web application response control. Here the search is performed in service information of the target system's response. The information that could appear only in case of incorrectly processed output is searched for. When this data is detected in the web application's response, the whole response is considered dangerous and is not carried through the WAF;

✧ signature analysis. A signature group is created in the WAF's internal storages. Each signature describes a case of a target system's manipulation attempt. If a sample in traffic contains the signature's data, the corresponding query is considered illegitimate.

The other threat is cross-site scripting (XSS) attempts. This becomes possible if a web application's response uses client-provided data without doing proper checks on this data. XSS allows the malefactor to steal client session identifiers, make web page defaces and re-route clients to arbitrary information resources. To detect XSS, the following techniques are applied:

● tokenization. Using a finite automaton, the query is parsed in order to search for declarative programming language tokens. If tokens valid for a programming language syntax are found, the current client query is declared potentially dangerous;

● content security policy (CSP) integration. A comparably new approach in information security: the CSP header defines for each web application response the possible resource sources which can be used to construct the page being displayed by the client's browser. The main difficulty: complexity of manual description of CSP rules. Some WAFs have techniques for CSP rules to be auto-created;

● response analysis. The response's content is matched to the data received from the client. If the data matches for the query-reply pair, the data transaction is declared illegitimate;

● web application response is injected with special Javascript code intended for page display control in the client's browser. This technique is most effective at detecting DOM-based XSS attempts;

● signature analysis. A signature group is created in the WAF's internal storage. Each signature describes a case of an XSS attempt. If a sample in the traffic contains the signature's data, the corresponding query is considered illegitimate.

5. User-defined rules of detecting illegitimate queries

WAF is an information security tool that is used «on top» of the protected server. It has a wide potential of features to analyze queries that go through it. Its capabilities, therefore, are:

✦ decryption;

✦ normalization;

✦ parsing;

✦ session control;

✦ traffic inspection;

✦ security policy checks;

✦ data leakage check.

These capabilities could be applied not only within the hard-coded mechanisms of WAF, but are presented to the information security administrator to form new, user-defined security rules. This could be useful when new security instructions are introduced in the information systems. Another case is adding new rules to prevent vulnerabilities detected during the information security auditing session. Moreover, when there is a need to define user logic from scratch and reflect it in website partitions access restriction, there is little to be done without user-defined security rules on WAF.

Generally, this mechanism's scenarios of usage are limited only by the toolset provided by WAF. Let's go through some examples of implementation.

The WAF product by F5 offers integrated programming language to define user rules. The scope of potential functions for this tool is huge, but it comes at a price: high competency requirements for the WAF administrator. Obvious difficulties could occur as well while implementing protection mechanisms requiring prompt intervention.

Toolset, offered by Imperva in their WAF, is sufficiently different from F5's. User rules here are generated by combining and correlation of a criteria set. The criteria amount totals about forty.

Doubtless, this kind of implementation provides far lower flexibility, particularly when speaking of bypassing traffic influence capabilities. However, these limitations are recouped by the relatively low threshold of competence for security administrators to manage the WAF.

6. «Denial of service» attack prevention

Ensuring availability of a protected resource is a task with the same importance level as data confidentiality

and integrity maintenance. Sometimes this task is even more important for systems sensitive to constant feedback to user (emergency control, payment processing, etc.). There is a prejudicial opinion that denial of service attacks should be handled at layers lower than the application level according to the ISO/OSI model. Nevertheless, WAF, acting at the application layer, offers interesting methods of prevention for this type of malicious activity.

WAF possesses a bot detection mechanism that is capable of telling whether there is a human operating a client machine or there is an automaton generating queries directed at a protected resource. Blocking automated machines hinders botnets' denial of service attack participation. This is reached through injection of a special javascript in the web application's replies to the client. The client must answer an easy question for WAF to draw a conclusion if there is a human-operated remote machine, or not.

Let's walk through other denial of service attack prevention mechanisms available in WAFs. First of all, WAF detects nodes involved in the attack. After this, reaction measures are applied to these nodes. A denial of service attack is defined based on queries per second or time required by the web-server to reply.

A query volume control mechanism works as follows. The last minute's statistics on number of queries are compared to the last 5 minutes' statistics. If the first value is higher than 5 times the second value (or if the first value has reached a previously defined level), active reaction measure performing mechanisms are enabled. Here, both statistics for every URL and IP-address are taken into account.

Turning to the web application time-to-reply controlling mechanism, it works the same way as the query volume control mechanism. Here the average delay for a query is analyzed and its ratio against 5-minute historic time interval is monitored.

As a result of any of these mechanisms' work, the WAF can enable one of the following attack response measures, thereby decreasing the attack's success chances:

✧ replies of web application are injected with a «proof of reply processing» Javascript-task, thus, slowing legal and illegal query sources. Nevertheless, we leave an availability window for an ensured number of users;

✧ the user and web-application interaction process is interrupted with a popup window containing «captcha»-trial. After the trial is completed, the session is considered to be with a «human»;

✧ bandwidth limitations for clients sending queries to the information resource being attacked are applied.

7. Integration of the information security landscape

An information security solution's efficiency is multiplied if different security software and hardware tools are interconnected. It is therefore important that products like WAF have wide integration capabilities to «understand» other security products and solutions and use data, generated by them to enhance its own security functions. Nowadays WAFs can be combined with the following system and services:

✦ vulnerability scanners;

✦ security information and event management systems;

✦ reputation services;

✦ fraud prevention services.

It seems that the most fruitful connection here is a vulnerability scanner data exchange. A function called «virtual patching» is implemented through this data exchange. It automates application security control: a scanner uncovers vulnerabilities and makes a report. Based on the information in the report, the WAF forms rules to block activity aimed at exploiting the newly-found vulnerability.

Information security incident control systems are a key point of information security efficacy for many large businesses. Using WAF, it becomes possible to take into account and correlate the events generated by web applications to events from other information and security systems.

Reputation services are specialized in revealing suspicious IP-addresses among those from the global IP-address pool. The database contains addresses of TOR endpoints, anonymous proxy servers, phishing and spam-generating nodes. The base also holds messages from the community participants on addresses violating their own security policies.

Fraud prevention systems are used by WAF to ensure that the client-server data exchange is not interfered with by third parties and that the client is not subject to a malware attack. WAF could use fraud prevention services to check that remote clients are «clean» and therefore raise the overall security level and lower the risks of scam and fraud operations.

Conclusion

Web application firewalls are a new product on the market in relation to traditional firewalls and anti-virus software. We have indicated the most-important features WAFs provide to information security engineers. With many vendors offering different WAF solutions on a broad market, understanding the WAF core functions and mechanisms is crucial to anyone who wants to build an even and balanced information security landscape in a company. Apart from these core functions and algorithm implementations, every solution possesses additional capabilities and offers new methods of ensuring web application information security. We intend to cover these methods and make a comparison of the most advanced products on the WAF market in our next article. ∎

References

1. Russian Federal Agency on Technical Regulating and Metrology (2013) *Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya bezopasnosti. Kriterii otsenki bezopasnosti informatsionnykh tekhnologii. Chast' 3. Komponenty doveriya k bezopasnosti* [Information technology. Security techniques. Evaluation criteria for IT security. Part 3: Security assurance components]. Moscow: Standardinform (in Russian).
2. Russian Federal Agency on Technical Regulating and Metrology (2008) *Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya bezopasnosti. Sistemy menedzhmenta informatsionnoi bezopasnosti. Trebovaniya* [Information technology. Security techniques. Information security management systems. Requirements]. Moscow: Standardinform (in Russian).
3. Bank of Russia (2014) *Rekomendatsii v oblasti standartizatsii Banka Rossii. Obespechenie informatsionnoi bezopasnosti organizatsii bankovskoi sistemy Rossiiskoi federatsii. Obespechenie informatsionnoi bezopasnosti na stadiyakh zhiznennogo tsikla avtomatizirovannykh bankovskikh system* [Bank of Russia standards recommendations. Russian Federation's Bank system organizations' information security. Information security maintenance in automated bank systems lifecycle] (electronic resource). Available at: http://www.cbr.ru/credit/Gubzi_docs/rs-26-14.pdf (accessed 29 September 2015) (in Russian).
4. Payment Card Industry (PCI) (2015) *Payment application data security standard. Requirements and security assessment procedures. Version 3.1* (electronic resource). Available at: https://www.pcisecuritystandards.org/documents/PA-DSS_v3-1.pdf (accessed 29 September 2015).
5. *Cisco Secure Development Lifecycle (SDL)* (electronic resource). Available at: http://www.cisco.com/web/about/security/cspo/csdl/index.html (accessed 29 September 2015).
6. *Microsoft Security Development Lifecycle* (electronic resource). Available at: http://www.microsoft.com/security/sdl/default.aspx (accessed 29 September 2015).
7. Beybutov E.R. (2015) *Obzor rynka zashchity web-prilozhenii (WAF) v Rossii i v mire* [Web Application Firewall (WAF) market review in Russia and over the world] (electronic resource). Available at: http://www.anti-malware.ru/reviews/web_application_firewall_market_overview_russia (accessed 29 September 2015) (in Russian).
8. *Exploit database.* Exploits for web applications (electronic resource). Available at: https://www.exploit-db.com/webapps/ (accessed 29 September 2015).
9. *Wallarm blog* (electronic resource). Available at: http://blog.wallarm.com (accessed 29 September 2015).

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ РЕСУРСОВ С ПОМОЩЬЮ МЕЖСЕТЕВЫХ ЭКРАНОВ ДЛЯ ВЕБ-ПРИЛОЖЕНИЙ

П.А. БАРАНОВ
кандидат технических наук, доцент кафедры инноваций и бизнеса в сфере информационных технологий, Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: pbaranov@hse.ru

Э.Р. БЕЙБУТОВ
студент магистратуры, факультет бизнеса и менеджмента, Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: eldar.beybutov@gmail.com

Данная работа содержит обзор основных технологий, реализуемых в относительно новых для рынка решений информационной безопасности продуктах — межсетевых экранах веб-приложений (веб-экранах). Веб-приложения представляют собой удобный и широко используемый способ предоставления доступа удаленных пользователей к корпоративным информационным ресурсам. Однако, сервер (серверы) веб-приложений может стать единой точкой отказа, таким образом, прерывая доступ легитимных пользователей к ресурсам информационной инфраструктуры организации. С целью создания инструмента противодействия злонамеренным попыткам доступа к защищаемым информационным ресурсам предприятия был разработан новый класс решений по защите информации — межсетевые экраны уровня приложений (веб-приложений).

Веб-экраны работают на седьмом уровне модели ISO/OSI и представляют собой туннель, контролирующий и, при необходимости, модифицирующий трафик, направленный к серверу веб-приложений и от него. Для обеспечения эффективного мониторинга и обработки трафика с целью выявления злонамеренных действий веб-экраны снабжаются различными механизмами контроля «нормальности» передаваемых в рамках сессии связи данных. Такие механизмы включают проверки соответствия правилам протокола формирования сообщений, техники машинного обучения и статистические подходы, анализ сигнатур в проходящем трафике, а также более узконаправленные средства, такие как предотвращение реализации атак типа «отказ в обслуживании», инъекции XSS и CRRF-атаки. В условиях неоднородности политик информационной безопасности на предприятиях и невозможности охватить все допустимые ограничения, налагаемые правилами обмена данными, несомненно актуальной является возможность добавления пользовательских правил к тем, которые уже были реализованы производителем. Веб-экраны обладают инструментариями, позволяющими создавать правила разной степени подробности.

Данное исследование основано на широком практическом опыте интеграции веб-экранов в существующие ландшафты информационной безопасности, их администрирования и настройки. С целью иллюстрации возможностей применения фильтрующих механизмов приводятся примеры их реализации в продуктах от ведущих игроков рынка решений информационной безопасности.

Литература

1. ГОСТ Р ИСО/МЭК 15408-3-2013. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 3. Компоненты доверия к безопасности. М.: Стандартинформ, 2013. 267 с.
2. ГОСТ Р ИСО/МЭК 27001-2006. Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования. М.: Стандартинформ, 2008. 23 с.
3. РС БР ИББС-2.6-2014. Рекомендации в области стандартизации Банка России. Обеспечение информационной безопасности организаций банковской системы Российской федерации. Обеспечение информационной безопасности на стадиях жизненного цикла автоматизированных банковских систем // Банк России. [Электронный ресурс]: http://www.cbr.ru/credit/Gubzi_docs/rs-26-14.pdf (дата обращения 29.09.2015).
4. Payment application data security standard. Requirements and security assessment procedures. Version 3.1. // Payment Card Industry (PCI). [Электронный ресурс]: https://www.pcisecuritystandards.org/documents/PA-DSS_v3-1.pdf (дата обращения 29.09.2015).
5. Cisco Secure Development Lifecycle (SDL). [Электронный ресурс]: http://www.cisco.com/web/about/security/cspo/csdl/index.html (дата обращения 29.09.2015).
6. Microsoft Security Development Lifecycle. [Электронный ресурс]: http://www.microsoft.com/security/sdl/default.aspx (дата обращения 29.09.2015).
7. Бейбутов Э.Р. Обзор рынка защиты веб-приложений (WAF) в России и в мире [Электронный ресурс]: http://www.anti-malware.ru/reviews/web_application_firewall_market_overview_russia (дата обращения 29.09.2015).
8. Exploit database. Exploits for web applications. [Электронный ресурс]: https://www.exploit-db.com/webapps/ (дата обращения 29.09.2015).
9. Wallarm blog. [Электронный ресурс]: http://blog.wallarm.com (дата обращения 29.09.2015).

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 представляется 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 − вверху 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 данным 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 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.

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.