

International Conference
e-Management
&
Business Intelligence

01-05 July 2007, Varna



P R O C E E D I N G S

ITHEA

SOFIA, 2007

Kr. Markov, Kr. Ivanova (Ed.)

Proceedings of the International Conference "e-Management & Business Intelligence".

Varna, 2007.

Sofia, Institute of Information Theories and Applications FOI ITHEA – 2007

First edition

The eM&BI conference is devoted to discussion of current research and applications regarding the basic directions of business informatics: e-management and business intelligence. eM&BI is complemented with Summer School on Business Informatics.

Printed in Bulgaria by Institute of Information Theories and Applications FOI ITHEA

Sofia -1090, P.O. Box 775

e-mail: info@foibg.com

www.foibg.com

All rights reserved.

© 2007 Krassimir Markov, Krassimira Ivanova - Editors

© 2007 Institute of Information Theories and Applications FOI ITHEA - Publisher

© 2007 For all authors in the issue

ISSN 1313-1141 (paperback) ISSN 1313-1184 (CD) ISSN 1313-1192 (online)

FSBN: 978-954-16-2016-8 (paperback) 978-954-16-2017-5 (CD) 978-954-16-2018-2 (Online)

PREFACE

The International Conference “**e-Management & Business Intelligence**” (**eM&BI 2007**) is organized as a part of “ITA 2007 - Joint International Scientific Events on Information Theories and Applications”.

ITA 2007 as well as the eM&BI 2007 is supported by

International Journal on Information Theories and Applications (IJ ITA)
and
International Journal on Information Technologies and Knowledge (IJ ITK)

eM&BI 2007 is dedicated to:

- 60th Anniversary of the Institute of Mathematics and Informatics of Bulgarian Academy of Sciences;
- 15th Anniversary of the Association of Developers and Users of Intelligent Systems (Ukraine);
- 10th Anniversary of the Association for Development of the Information Society (Bulgaria).

The eM&BI conference is devoted to discussion of current research and applications regarding the basic directions of business informatics: e-management and business intelligence.

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

Let us thank the Program Committee of the conference for referring the submitted papers. Special thanks to prof. Viktor Gladun, prof. Alexey Voloshin, prof. Koen Vanhoof and prof. Avram Eskenazi.

eM&BI 2007 Proceedings has been edited in the *Institute of Information Theories and Applications FOI ITHEA* in collaboration with the leading researchers from *Institute of Cybernetics “V.M.Glushkov”, NASU (Ukraine), Kiev University “T.Shevchenko” (Ukraine), Institute of Mathematics and Informatics, BAS (Bulgaria), Institute of Information Technologies, BAS (Bulgaria)*.

The eM&BI 2007 Conference found the best support in the work of Organizing Committee Chairman Ilia Mitov.

To all participants of eM&BI 2007 we wish fruitful contacts during the conference days and efficient work for preparing the high quality papers to be published in the International Journal on Information Theories and Applications or the International Journal on Information Technologies and Applications.

Varna, July 2007

Kr Markov, Kr. Ivanova

eM&BI 2007 has been organized by:

- Institute of Information Theories and Applications FOI ITHEA (Bulgaria)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Information Technologies, BAS (Bulgaria)
- Taras Shevchenko National University of Kiev (Ukraine)
- Kharkiv National University of Radio Electronics (Ukraine)
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Association for Development of the Information Society (Bulgaria)
- International Journal "Information Theories and Applications"
- International Journal "Information Technologies and Knowledge"

Program Committee

Koen Vanhoof (Belgium)
 Krassimir Markov (Bulgaria)
 Alexander Lounev (Russia)
 Irina Petrova (Russia)

Alexander Kuzemin (Ukraine)	Levon Aslanyan (Armenia)
Alexander Palagin (Ukraine)	Luis de Mingo (Spain)
Alexey Voloshin (Ukraine)	Mikhail Bondarenko (Ukraine)
Avram Eskenazi (Bulgaria)	Nikolay Lyutov (Bulgaria)
Anna Kantcheva (Bulgaria)	Peter Stanchev (Bulgaria)
Ekaterina Solovieva (Ukraine)	Radoslav Pavlov (Bulgaria)
Ilia Mitov (Bulgaria)	Sregey Turkov (Russia)
Ivan Popchev (Bulgaria)	Stefan Dodunekov (Bulgaria)
Jeanne Schreurs (Belgium)	Stoyan Poryazov (Bulgaria)
Juan Castellanos (Spain)	Tatiana Atanasova (Bulgaria)
Krassimira Ivanova (Bulgaria)	Vitaliy Lozovskiy (Ukraine)
Laura Ciocoiu (Romania)	Vladimir Ryazanov (Russia)

Organizing Committee:

Ilia Mitov	Krassimira Ivanova	Emilia Saranova	Stoyan Poryazov
Todorka Kovacheva	Tsvetanka Kovacheva	Valeria Dimitrova	Vera Markova

Topics of interest include, but are not limited to:

- e-Management in Governmental and Municipal Structures: Models, Systems, e-Government, etc. ;
- Business Informatics: e-Management, e-Finance, e-Commerce, e-Banking;
- Management and Economics of Education;
- Communication Technologies and Mobile Systems in e-Management;
- Business Intelligence: Methodology, Tools and Technologies, Analytics and Statistics;
- Competitive Intelligence;
- Corporative e-Culture.

Official languages of the conference are English and Russian.

General sponsor of the eM&BI 2007 is **FOI BULGARIA** (www.foibg.com).

TABLE OF CONTENTS

<i>Preface</i>	3
<i>Table of Contents</i>	5
<i>Index of Authors</i>	6

Papers

Systemological Bases of Management Consulting <i>Mikhail Bondarenko, Ekaterina Solovyova, Dmitriy Elchaninov</i>	7
Hardware-based and software-based security in Digital Rights Management solutions <i>Maria Nickolova, Eugene Nickolov</i>	12
The Strategical Impact of Knowledge Markets and Knowledge Management for Small and Medium Enterprises <i>Todorka Kovacheva</i>	17
An Approach to Representing the Process of Information Business Modeling <i>Nadezhda Filipova, Filcho Filipov</i>	20
The Role of DBMS in Analytical Processes of the Logistic of Stock Reserves <i>Julian Vasilev</i>	26
Bivirtual Organization as a Queuing System <i>Tetiana Palonna, Iurii Palonnyi</i>	29
Identifying Business Components from Business Model: a Method Based on Feature Matching <i>Meng Fanchao, Zhan Dechen, Xu Xiaofei</i>	33
Application of Intellectual Technologies in Profit Control <i>Inna Khmelnitskaya, Oleg Kosenkov</i>	41
Development of Database for Distributed Information Measurement and Control System <i>Sergey Kiprushkin, Sergey Kurskov, Vadim Semin</i>	48
Expert System for Decision-Making Problem in Economics <i>Aygun Alasgarova, Leyla Muradkhanli</i>	52
Автоматизированная система принятия решений и объектного мониторинга мероприятий, проводимых на территории региона <i>Ирина Петрова, Марина Зайцева, Оксана Щербинина, Евгений Эрман, Николай Тюрин</i>	55
Об одном подходе к проектированию интерфейса автоматизированной документальной системы <i>Арсений Баканов</i>	62
Разработка комбинированного метода для проектирования систем документооборота <i>Нина Баканова</i>	65
Метод анализа иерархий как инструмент обоснования бизнес-решений <i>Михаил Плаксин</i>	68
Выявление несоответствия в модели гена информационной системы <i>Виктор Левыкин, Максим Евланов</i>	75
Синергические аспекты процесса поиска равновесной цены объекта недвижимости <i>Марина Чичужко, Валерий Тазетдинов</i>	78

Lecture notes for Summer School

Concepts and Good Practices in Reporting, Business Intelligence and Datawarehousing <i>Koen Vanhoof</i>	83
The Entrepreneurship in Information and Telecommunication Technologies – an Experience from one course in Berkeley <i>Neli Maneva</i>	97
Электронный рынок знаний <i>Красимир Марков, Красимира Иванова, Илия Митов</i>	98

About

<i>60th Anniversary of Institute of Mathematics and Informatics, Bulgarian Academy of Science</i>	109
--	-----

INDEX OF AUTHORS

Aygun Alasgarova	52	Neli Maneva	97
Arsenii Bakanov	62	Krassimir Markov	98
Nina Bakanova	65	Ilia Mitov	98
Mikhail Bondarenko	7	Leyla Muradkhanli	52
Marina Chichuzhko	78	Eugene Nickolov	12
Zhan Dechen	33	Maria Nickolova	12
Dmitriy Elchaninov	7	Tetiana Palonna	29
Evgeniy Erman	55	Iurii Palonnyi	29
Maxim Evlanov	75	Irina Petrova	55
Meng Fanchao	33	Mikhail Plaxin	68
Filcho Filipov	20	Vadim Semin	48
Nadezhda Filipova	20	Oxana Shcherbinina	55
Krassimira Ivanova	98	Ekaterina Solovyova	7
Inna Khmel'nitskaya	41	Valerii Tazetdinov	78
Sergey Kiprushkin	48	Nikolay Tyurin	55
Oleg Kosenkov	41	Koen Vanhoof	83
Todorka Kovacheva	17	Julian Vasilev	26
Sergey Kurskov	48	Xu Xiaofei	33
Viktor Levykin	75	Marina Zaitseva	55

SYSTEMOLOGICAL BASES OF MANAGEMENT CONSULTING

Mikhail Bondarenko, Ekaterina Solovyova, Dmitriy Elchaninov

Abstract: *The problem of management consulting for sustainable development organization support is discussed. The problem is formally described by means of systemological terms. The mathematical problem solving is considered. Practical use of the obtained results is outlined.*

Keywords: *systemology, competitive intelligence, management consulting, sustainable development.*

ACM Classification Keywords: *H. Information Systems - H.1 Models and Principles - H.1.1 Systems and Information Theory - General systems theory*

Introduction

At the beginning we will discuss the sustainable development firm problem. Then the problem will formally described by means of systemological terms. The end of the paper will be devoted to mathematical problem solving. In conclusion we will outline practical use of the obtained results.

Management Consulting for Sustainable Development

According to [Greiner and Metzger, 1983]: management consulting is an advisory service contracted for and provided to organizations by specially trained and qualified persons who assist, in an objective and independent manner, the client organization to identify **management problems**, analyze such problems, recommend solutions to these problems, and help, when requested, in the implementation of solutions.

Staffan Canback defines management consultants as those who provide general management advice within a **strategic, organizational or operational context** [Canback, 1998].

The contexts correspond with three management levels: strategic, tactical and operational. Strategic managers focus on long-term issues and emphasize the **survival, growth and overall effectiveness of the organization**. Tactical managers are responsible for translating the general goals and plans developed by strategic managers into more specific objectives and activities. Operational managers are directly involved with nonmanagement employees, implementing the specific plans developed with tactical managers. [Bateman and Snell, 1996].

Management is about helping a firm **survive and win** in competition with other companies. The firm gains competitive advantage by adopting **management approaches that satisfy people** (both inside and outside the firm) through cost competitiveness, high-quality products, speed and innovation [Lawler, 1992].

The aim of many companies is to jointly achieve the goals of economic growth and environmental quality in the long run by striving for **sustainable growth**. Sustainable growth is economic growth and development that meets the organization present needs without harming the ability of future generations to meet their needs [Rice, 1993].

The first thing managers can do to better understand environmental issues in their companies is to engage in systems thinking. Managers operate in organizations. An organization is a managed system designed and operated to achieve a specific set of objectives. Management scholars during the 1950s stepped back from the details of the organization to attempt to understand it as a whole system. These efforts were based on a general scientific approach called **systems theory** [Bertalanffy, 1972]. Business research is largely supported by business organizations that hope to achieve a competitive advantage [Cooper and Schindler, 2001].

In [Bossel, 1999] systems theory is used to identify the vital aspects of **sustainable development** and relevant indicators. Much work has been devoted to developing quantitative indicators of sustainable development [Parris and Kates, 2003], [Segnestam, 2002], [Harris, 2000]. A great deal of literature concerned with understanding the core principles of sustainable development [Scottish Executive, 2006].

The aim of this paper is to show the efficiency of systemology [Melnikov, 1988] as a new concept for a system approach to solving **sustainable development** problems. The systemology was successfully used to solve a natural classification problem [Bondarenko et al, 2001] and business-systems modeling [Matorin and El'chaninov, 2002]. We hope that our results will be useful for sustainable business-systems modeling in management consulting process.

Properties of a System

In connection with the fact that the terminology of the suggested systemological approach is not widely known, we will give a list of terms needed to understand the essentials of the present investigation [Melnikov, 1988], [Bondarenko et al, 1996].

System - an object the **properties** of which are determined by a function, which amounts to maintaining certain properties of an object at a higher level. This object is a **supersystem** in relation to the object (system) under consideration. Substance of a system - elements or **components** of the system, usually considered as subsystems. **Structure** of a system - the scheme of relations and interactions of a system's substance. **Property of a system** (valence) - **the ability to maintain** (in certain conditions) **relations of one type** and to prevent realization of relations of other types. Functional property of a system - a property that a system must possess in order to perform its functions; the ability to maintain relations (flows) on the basis of which interactions that are important for the supersystem occur between the system and surrounding systems. **Extensional** valence - a **property** realized in the form of a relation of the corresponding quality and constituting one of the varieties of reality. Free valence - a **property** only as an ability, not manifested in an existing relation and constituting one of the varieties of possibility (weak: **potential**, strong: **intentional**). External determinant of a system - the main reason for formation of a system: **the supersystem's functional need** for certain interactions of the system under consideration with other (surrounding) systems of this supersystem, which dictates the choice of the system's determinant.

Ports of a System

A port of a system S is an input or an output of the system S . By means of ports the system S maintains relations with another systems. In other words system's ports correspond to system's properties. Extensional, intentional and potential ports of the system S are represented as figure 1 shows.

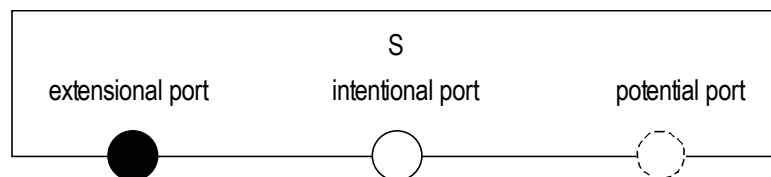


Fig. 1

An input port In is equal to an output port Out if the type of the port In is equal to the type of the port Out and vice versa. It means that if a system S_i has an output port $Out_{i,x}$, which is equal to an input port $In_{j,y}$ of a system S_j , then systems S_i and S_j can be connected with each other as figure 2 shows.

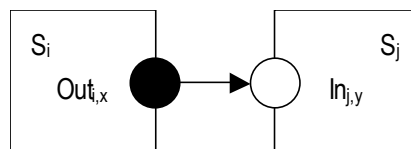


Fig. 2

Let $In(S)$ is a set of the input ports of a system S and $Out(S)$ is a set of the output ports of a system S .

An input port In_i of a system S is connected with an output port Out_j of the system S by means of component sequence $\{S_q\}$ (see fig. 3) if:

1. $In_i \in In(S_1)$;
2. $Out(S_q) \cap In(S_{q+1}) \neq \emptyset$, ($q=1, \dots, Q-1$);
3. $Out_j \in Out(S_Q)$.

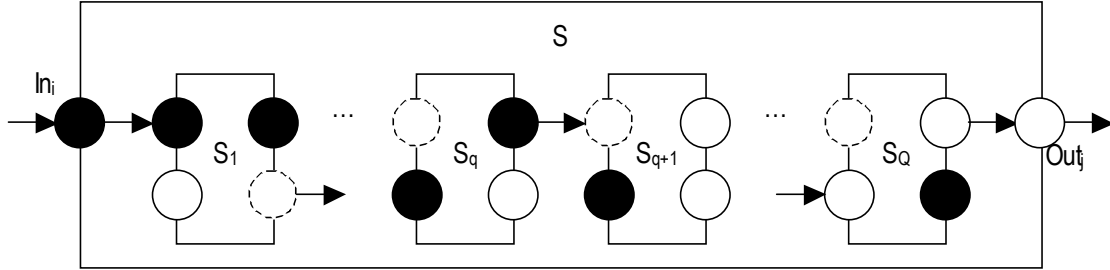


Fig. 3

Let this connection is denoted as follows: $\langle In_i, S_1, \dots, S_q, S_{q+1}, \dots, S_Q, Out_j \rangle$.

A subset $B=\{B_{ij}\}$ of the set of components $C=\{C_k\}$ is satisfied to a system S if each ports from the set $In(S)$ is connected with some port from the set $Out(S)$ by means of some subset of the subset B and vice versa, i.e.:

1. $\forall In_i \in In(S) \exists Out_j \in Out(S) \exists D_i = \{D_{i,v}\} \subset B: \langle In_i, D_{i,1}, \dots, D_{i,v}, D_{i,v+1}, \dots, D_{i,v(j)}, Out_j \rangle$;
2. $\forall Out_j \in Out(S) \exists In_i \in In(S) \exists E_j = \{E_{j,w}\} \subset B: \langle In_i, E_{j,1}, \dots, E_{j,w}, E_{j,w+1}, \dots, E_{j,w(j)}, Out_j \rangle$.

Formal Statement of Problem

It is given:

1. External determinant of a system S - a set $\{T_m\}$ of systems with free ports $Out(T_m)$ and a set $\{P_n\}$ of systems with free ports $In(P_n)$.
2. A set (library, package) of components $C=\{S_k\}$.

It is required to find a substance $B \subset C$ of a system S and a structure of a system S such that $\cup Out(T_m) \subset In(B)$ and $\cup In(P_n) \subset Out(B)$ (see fig. 4).

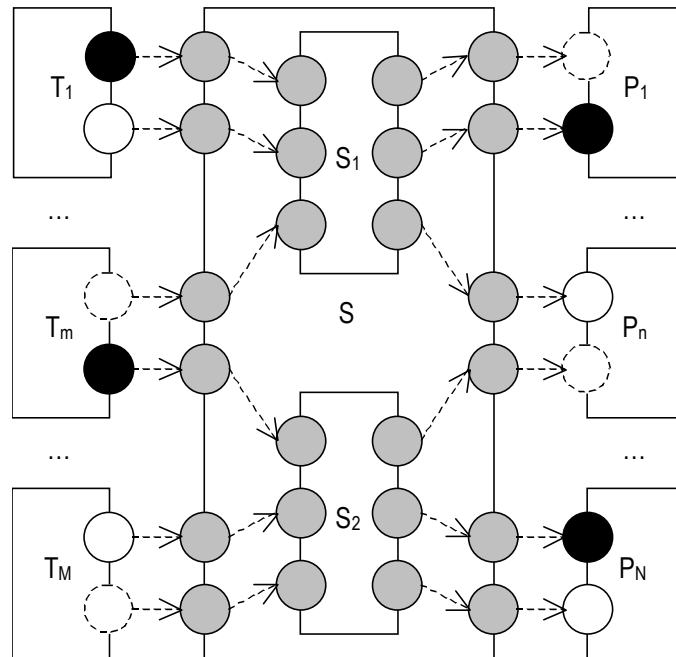


Fig. 4

Problem Solving

Lemma 1. The necessary conditions of being a set of components, which is satisfied to a system.

If a set of components $\{S_u\}$ is satisfied to a system S then $\text{In}(S) \subseteq \bigcup \text{In}(S_u)$ and $\text{Out}(S) \subseteq \bigcup \text{Out}(S_u)$.

The proof is illustrated by fig. 5.

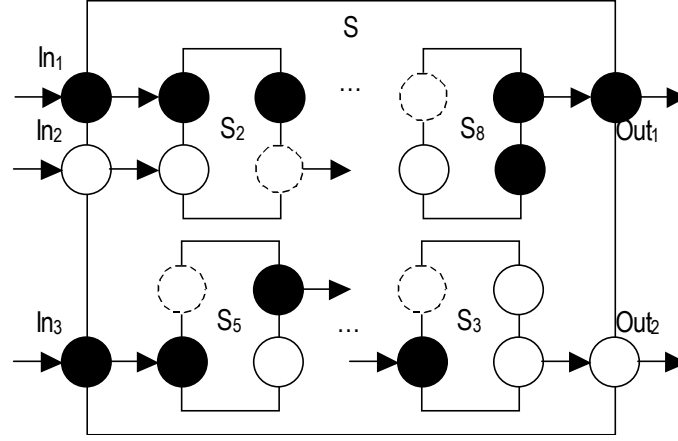


Fig. 5

Lemma 2. The sufficient conditions of excluding a component from a set of components, which is satisfied to a system.

If $\text{Out}(S_k) \cap \text{Out}(S) = \emptyset$ and $\text{Out}(S_k) \cap (\bigcup \text{In}(S_u)) = \emptyset$ or $\text{In}(S_k) \cap \text{In}(S) = \emptyset$ and $\text{In}(S_k) \cap (\bigcup \text{Out}(S_u)) = \emptyset$ then $S_k \notin \{S_u\}$, where the set of components $\{S_u\}$ is satisfied to the system S .

The proof is illustrated by fig. 6.

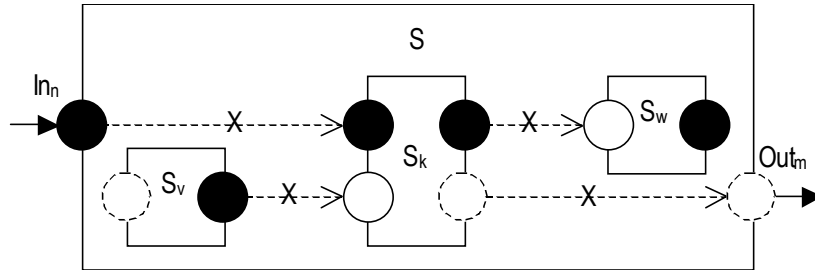


Fig. 6

Lemma 3. It is the consequence of lemma 2.

If C^* is a set of components, which is satisfied to the conditions of lemma 2, then any component S_u from a set $C \setminus C^* = \{S_u\}$ is such that:

1. $\text{In}(S) \cap \text{In}(S_u) \neq \emptyset$ or $\exists S_v \in \{S_u\}: \text{Out}(S_u) \cap \text{In}(S_v) \neq \emptyset$;
2. $\text{Out}(S) \cap \text{Out}(S_u) \neq \emptyset$ or $\exists S_w \in \{S_u\}: \text{Out}(S_u) \cap \text{In}(S_w) \neq \emptyset$.

The proof is trivial.

Theorem. The sufficient conditions of being a set of components, which is satisfied to a system.

If $\exists \{S^r\}: \bigcup S^r \subseteq C$, where $S^r = \{S_{r,u}\}$, are such that:

1. $\forall S_{1,v} \in S^1 \Rightarrow (\text{In}(S) \cap \text{In}(S_{1,v}) \neq \emptyset) \wedge (\text{In}(S) \subseteq \bigcup \text{In}(S_{1,v}))$;
2. $\forall S_{r,v} \in S^r \Rightarrow (\exists S_{r-1,v} \in S^{r-1}: \text{Out}(S_{r-1,v}) \cap \text{In}(S_{r,u}) \neq \emptyset) \wedge (\exists S_{r+1,w} \in S^{r+1}: \text{Out}(S_{r,u}) \cap \text{In}(S_{r+1,w}) \neq \emptyset), r=2, \dots, R-1$;
3. $\forall S_{R,u} \in S^R \Rightarrow (\text{Out}(S) \cap \text{Out}(S_{R,w}) \neq \emptyset) \wedge (\text{Out}(S) \subseteq \bigcup \text{Out}(S_{R,w}))$,

then the set of components $\bigcup S^r$ is satisfied to the system S .

The proof is left to the reader.

Conclusion

Using the obtained results, one can make sustainable organization modeling. Lemma 1 can be used for selection of component libraries. Lemma 2 is useful for exclusion of components, which are unfit for system modeling. Lemma 3 describes all possible connections of components, which can be used for sustainable system modeling. Finally, the theorem is useful for assembling of complex multi-tier system configuration. The results can be applied in powerful modeling tool for analyzing, documenting and understanding complex business processes.

Bibliography

- [Bateman and Snell, 1996] Bateman T.S., Snell S.A. Management: Building Competitive Advantage. Richard D. Irwin, 1996.
- [Bertalanffy, 1972] von Bertalanffy L. The History and Status of General Systems Theory. Academy of Management Journal, 1972, V.15, pp. 407-426.
- [Bondarenko et al, 1996] Bondarenko M.F., Matorin S.I., Solov'eva E.A. Analysis of Systemological Tools for Conceptual Modeling of Application Fields. Automatic Document and Mathematical Linguistics. Allerton Press, Inc., 1996, V.30, No.2, pp. 33-45.
- [Bondarenko et al, 2001] Bondarenko M.F., Elchaninov D.B., Solovyova E.A., Matorin S.I. Systemological and Mathematical Foundations of a Natural Classification. IJ ITA, 2001, V.8, No.3, pp. 151-157.
- [Bossel, 1999] Bossel H. Indicators for Sustainable Development: Theory, Method, Applications. A Report to the Balaton Group. International Institute for Sustainable Development: Winnipeg, Canada, 1999.
- [Canback, 1998] Canback S. The Logic of Management Consulting, Part 1. Journal of Management Consulting, 1998, V.10, No.2, pp.3-11.
- [Cooper and Schindler, 2001] Cooper D.R., Schindler P.S. Business Research Methods. New York: McGraw-Hill/Irwin, 2001.
- [Greiner and Metzger, 1983] Greiner L., Metzger R. Consulting to management. Englewood Cliffs, N.J.: Prentice-Hall, 1983.
- [Harris, 2000] Harris J.M. Basic Principles of Sustainable Development. Working Paper No. 00-04. Global Development and Environment Institute: Tufts University, USA, 2000.
- [Lawler, 1992] Lawler E.E. The Ultimate Advantage. San Francisco: Jossey-Bass, 1992.
- [Matorin and El'chaninov, 2002] Matorin S.I., El'chaninov D.B. Application of Pattern Theory in Formalization of Systemological UFO Analysis. Automatic Document and Mathematical Linguistics. Allerton Press, Inc., 2002, V.36, No.6, pp. 1-12.
- [Melnikov, 1988] Melnikov G.P. Systemology and Linguistic Aspects of Cybernetics. New York, Paris, Monterey, Tokyo, Melbourne: Gordon and Breach. 1988.
- [Parris and Kates, 2003] Parris T.M. and Kates R.W. Characterizing and Measuring Sustainable Development. Annual Review of Environment and Resources, 2003.
- [Rice, 1993] Rice F. Who Scores Best on the Environment. Fortune, July 26, 1993, pp. 114-122.
- [Scottish Executive, 2006] Sustainable Development: A Review of International Literature. Edinburgh: Scottish Executive, 2006. <http://www.scotland.gov.uk/Publications/2006/05/23091323/17>
- [Segnestam, 2002] Segnestam L. Indicators of Environment and Sustainable Development: Theories and Practical Experience. Environment Department Papers No. 89. The World Bank: Washington, DC, USA, 2002.

Authors' Information

Mikhail Bondarenko - Rector, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: rector@kture.kharkov.ua

Ekaterina Solovyova - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua

Dmitriy Elchaninov - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: elchaninov@kture.kharkov.ua

HARDWARE-BASED AND SOFTWARE-BASED SECURITY IN DIGITAL RIGHTS MANAGEMENT SOLUTIONS

Maria Nickolova, Eugene Nickolov

Abstract: *The main requirements to DRM platforms implementing effective user experience and strong security measures to prevent unauthorized use of content are discussed. Comparison of hardware-based and software-based platforms is made showing the general inherent advantages of hardware DRM solutions. Analysis and evaluation of the main flaws of hardware platforms are conducted, pointing out the possibilities to overcome them. The overview of the existing concepts for practical realization of hardware DRM protection reveals their advantages and disadvantages and the increasing demand for creation of multi-core architecture, which could assure an effective DRM protection without decreasing the user's freedom and importing risks for end system security.*

Keywords: *Security, DRM protection.*

ACM Classification Keywords: *D.4.6 Security and Protection.*

Introduction

Security design is one of the most challenging areas for system designers because it requires an extraordinary effort to build a system offering strong security features but not hindering the working process of users and being well accepted by them. This is particularly true as far as the compromise between the content owner's copyrights and the right of free access and exchange of information is concerned. The solution adopted in last decade is the digital rights management. Although most users don't agree with the use of DRM, it is of critical importance for authors, publishers and content providers - their business depends on the ability to control and to charge for access to their content.

Although the inherent insecurity of Internet, many upper-layer security protocols can be used to protect data during transmission but content is still at risk when it arrives at its destination. If the end device's boot process and critical information are not highly secure, the digital content can be stolen after the transmission and distributed without permission. This implies that end user devices must be built on a trusted platform and equipped with mechanisms for cryptographically validating the hardware environment and code signatures of downloaded software [1].

The DRM technologies allowing the protection of the content by access from unauthorized users could be divided into three groups: DRM implemented completely by software, DRM implemented completely by hardware, and the hybrid combinations of software and hardware. Certainly the most secure DRM is that which is implemented by hardware, the next most secure is the hybrid, and the least secure is via software.

Main requirements to DRM platforms

An effective DRM technology must provide a smooth and effective user experience for content use and in the same time must implement strong security measures to prevent unauthorized use of content [2]. The main requirements to it are:

1. It must ensure fully protected capabilities, which means the protection functions should be performed as part of the boot process. Otherwise during boot-up malicious software can easily hook the control functions and compromise system integrity. If end devices receive content over a network, such malicious software could be masked as a firmware upgrade or Trojan, or hidden using rootkits.

2. It must allow trusted integrity measurement and confirmation, that means the platform should own the capability to automatically check in real time during the boot all the new software and executable files in the system (certificates, digital signatures). Once this confirmation is done, the operating system loader can be started and the boot process proceeds as normal.

3. It must provide integrity reporting to notify the user about the results of the integrity measurement and possibly to prevent the user from playing back the DRM protected content in case of negative results from the integrity check.

Obviously these requirements could be implemented by hardware and/or software means.

Advantages of hardware-based DRM versus software-based

The analysis of the commercially available technologies for DRM protection shows two main reasons to use hardware-based security of the protected content: better overall robustness and improved user experience. The main benefits of the hardware-based security robustness are:

- Immunity from the inherent vulnerabilities and security holes of the used operating system. The security of all software applications is limited by the level of security provided by the underlying OS. Although the open and rich OS have bigger security challenges than a closed OS a hardware security module is an essential element to make the OS trustworthy.
- Impossibility to access, change or uninstall security features. Attacks to DRM protection often start by targeting the protection software - trying to uninstall it or stop its activity [3]. Obviously hardware-based DRM protection cannot be uninstalled as it is hard coded into the chips.
- Protected memory. Hardware-based DRM solutions manage the memory in a restricted manner and are able to prohibit access to it, providing better protection against attacks on the security mechanism. Software solutions use memory by the services of the operating system and several processes can access the same memory space simultaneously. Most OS provide some memory protection, but the safety of the memory space depends on the extent to which the operating system is robust and free of flaws. This is particularly important for the cryptographic algorithms which require the storage of the intermediate results during the execution of the cryptographic module. If the content of this temporary storage is exposed, the entire DRM system can be easily compromised.
- Better performance. The hardware DRM protection could be optimized for maximum security and operate independently, not degrading the performance of the computer or consuming its resources.
- Prevention of potential software conflicts. The software DRM protection is run on the same computer with many other security programs using together the same processor, memory, OS and other resources. This could provoke various conflicts resulting in poor performance and even in stopping the action of both DRM protection software and security programs.
- Secure Storage. Hardware-based DRM protection is able to better protect sensitive data, such as private keys. A software DRM implementation cannot prevent the exposure of keys and therefore they could be relatively easily compromised. Even very strong cryptographic algorithms could be easily compromised by an direct or indirect attack to their software implementation. Only a proper hardware implementation, to which countermeasures against known attacks are added, could protect the secrecy and the integrity of the DRM mechanism.
- True Random Number Generation. The software DRM technologies use pseudo-random numbers that decrease the security level of the DRM protection. As random numbers are used in DRM protection process

for the creation of temporary and special values and are part of challenge response authentication, the better the random number generator, the more secure the DRM implementation.

- Easier, faster and cheaper attacks to software DRM solutions. This is related to the security vulnerabilities, which are inherent for software modules and to the presence of many hackers who have enough time, knowledge and wish for breaking the relevant protection.
- Quick dissemination. The compromising of software DRM solutions by only one hacker becomes quickly available for general use. The publishing in Internet of correspondent methodology allows it to be used by a lot of end-users before the manufacturer could take measures to remove the vulnerabilities in the protection, and to bring severe damage to operators, content providers and manufacturers.
- Less susceptibility to reverse engineering. Hardware-based platforms are able to apply special measures that hide the data-dependent fluctuations in power consumption while software-based DRM solutions are more vulnerable to attacks based on power analysis.
- Most content applications like music, video and games require efficient and effective user experience which is the key factor for the success of consumer electronic devices and therefore for the acceptance of DRM by users. The main benefits of the hardware-based improvement of user experience are:
 - Superior performance in which user experience is prioritized without sacrificing security. Hardware solutions generally accelerate several times cryptographic functions (which are computation-intensive) in comparison with software solutions, making DRM security operations almost invisible for the end-user.
 - Optimization of CPU power and memory use. Although the computing power of modern processors increases constantly and should allow relatively fast handling of cryptographic functions, processors are designed mainly for new demanding applications such as video rendering and high quality graphics. Therefore software-based cryptographic operations are able to overload them and to worsen the user experience. There are some cryptographic operations (exchange of protocols with long keys, for example) that affect inadmissibly user experience.
- Improved power consumption and memory use. The use of hardware-based DRM platforms allows the CPU to operate at a lower clock rate, saving power which is particularly important for battery powered mobile devices. Additionally, software-based solutions require more memory (code size needs large buffers) which affect the speed and the quality of other applications.

Disadvantages of hardware-based DRM platforms

- Software modification or creation of new software on a computer with hardware-based DRM technology may require hardware changes that could be slow and expensive.
- The simple replacing of a peripheral device running protected content could cause a hardware-based DRM system to refuse to run software.
- Network cards replacement could make a computer unusable until other necessary hardware modifications are done and passwords are reauthorized. This process may require the cooperation of several vendors.
- If DRM protection is compromised reinstalling is impossible.
- Manufacturers of hardware-based DRM are not able to warrant that DRM agents or their hardware assistants will not cause or help any safety or security failures.
- More difficult implementation of extended usability. Software DRM implementations facilitate the making of the licensed content usable by a user anywhere in his personal network (local hard drive, media center,

iPOD, cell phone, home entertainment center or burning to a CD), for the hardware-based it's more difficult and expensive.

- Higher security in hardware-based DRM solutions means higher costs, less interoperability, longer development cycles and potentially shorter market life.
- Limited flexibility. It's difficult to make hardware DEM systems open to new uses, new business models, or new rights created by content owners.

It is clear that these flaws could be easily overcome and that only a hardware-based DRM implementation or a hybrid hardware/software solution could address all required security challenges while allowing seamless user experience [4].

Approaches for implementing hardware-based DRM

Two main concepts have been developed by now: trusted system concept and multi-core concept.

Trusted system concept

The Trusted Computing Group (TCG), successor of the Trusted Computing Platform Alliance (TCPA), is an initiative led by AMD, IBM, Intel, Hewlett-Packard, Microsoft, Sony, and Sun Microsystems. Its aim is to develop and promote open, vendor-neutral, industry standard specifications for trusted computing building blocks and software interfaces across multiple platforms [5]. The new principles in the TCG architecture expand the range of entities that are able to use TCG features as a trust basis. These entities could include not only the direct user of the platform and the owner but also some remote entities wishing to interact with this platform. The TCG architecture introduces the mechanism of remote attestation which allows remote third parties to ask a platform for details of its current software state. On the basis of the attestation made, third parties can decide to consider the platform's configuration as trustworthy or not. If correctly implemented, this kind of remote attestation could become an important feature for DRM clients on open platforms as it may help a content provider when he makes a decision about the reliability of the client before the content is actually provided. What makes TC technology especially attractive for implementing DRM is its ability to enforce usage policies. Once their security conditions are violated, TC systems stop working. Since their security conditions are built as a "chain of trust" [6] containing hardware-locked keys and certificates from trusted third parties, they are hard to modify, at least much harder than software-based systems. If a DRM solution relies on a trusted system, it is easy to implement a hard-to-break usage rights management chosen by content owners. TC technology is not necessary or sufficient to implement DRM but it can make implementing DRM easier and cheaper. An example of such a realization is the Intel Wireless Trusted Platform with the Certicom Security Architecture software. In this technology a special trusted platform module is built directly into the processor and provides secure key and password storage and protection. First, a secure boot process authenticates the hardware platform and the security architecture authentication module, then the module runs DRM applications and allows the users to access DRM protected content. The security architecture requires decoding the keys using information stored in secure hardware, to be able to access the content, after what these keys are used to decrypt and use the content, but only on the specific device. The encrypted content is not locked to this device, because another user is not able to use the content without paying to the content provider for having access to the rules for the content use [7].

Multi-core concept

Intel's Hyper-Threading technology allows parallel processing at thread-level on a single-core processor by sharing the processor's resources. In Intel multi-core processor, each thread is processed independently by a separate dedicated processor, which allows full parallel execution at hardware-level and software-level and is very suitable for DRM applications [8].

In 2005 Intel embedded DRM capabilities within its dual-core processor Pentium D and allowed (theoretically) copyright holders to prevent unauthorized use and distribution of DRM protected materials [9]. But some functional problems with the distribution of jobs in the cores and in the chip-set when both cores are enabled

caused applications to crash or hang and finally made hardware DRM capabilities unusable for real protection of content. The next stage in the implementation of hardware-based DRM in Intel's products was Lenovo's ThinkPad model, launched in 2006. It uses a combination of fingerprint sensor, trusted platform module chip and special software from Microsoft and Adobe to control access and distribution only of PDF documents.

AMD also planned to incorporate DRM into future GPUs by blocking the access to the frame buffer and allowing access only to certain software from certain vendors but these plans didn't involve AMD multi-core processors because of the complexity of problems in sharing and synchronizing DRM-related actions.

In 2006 IBM announced the technology Secure Blue intended for use in digital media players, electronic organizers, mobile phones, computers and devices where data is encrypted and decrypted as it runs through the processor and maintained encrypted in the device's RAM. Secure Blue requires a few circuits to be added to any processor design in order to enforce strict access controls at the hardware level.

Conclusion

It is obvious that DRM is becoming an integrated part of any copyright protected intellectual product in digital form and therefore DRM protection should be implemented in hardware and/or software assuring highest stability and performance as well as the best copyright protection possible. Different adopted solutions have many advantages and disadvantages but clearly show that it is impossible to realize well working solutions based only on software tools. Hardware-based platforms, especially those using multi-core processors demonstrate really promising features by improving user experience along with the robustness of DRM protection. Technology from a hardware standpoint is already in place, thanks to the efforts of various chipset manufacturers who have driven an evolution to support the benefits of parallel processing. Now research must be conducted to develop suitable multi-CPU architectures and multithreaded software that will guarantee the building of the perfect DRM system – fast, flawless and cheap - that can be neither broken nor avoided.

Bibliography

- [1] Biddle, P., England, P., Peinado, M. and Willman, B. (2003). The Darknet and the future of content protection. In Digital Rights Management-Technological, Economic, Legal and Political Aspects. LNCS 2770, Springer.
- [2] CEN/ISSS, (2003). Digital Rights Management Report <http://europa.eu.int/comm/enterprise/ict/policy/doc/drm.pdf>.
- [3] M. Nickolova, E. Nickolov, Verification and Application of Conceptual Model and Security Requirements on Practical DRM Systems in E-Learning. In: First International Workshop "Cyber Security" - CS 2006.
- [4] Hauser, T. and Wenz, C. (2003): DRM Under Attack: Weaknesses in Existing Systems. In Digital Rights Management-Technological, Economic, Legal and Political Aspects. LNCS 2770, Springer.
- [5] Peinado, M. Chen, Y. et al. (2004), NGSCB: A Trusted Open System. In Proceedings of 9th Australasian Conference on Information Security and Privacy ACISP, Sydney, Australia, July 13-15.
- [6] Smith, S.W. (2005): Trusted computing platforms: Design and applications. Berlin, Heidelberg, New York: Springer.
- [7] Dornan, A. (2006): Yes, trusted computing is used for DRM; Information Week, 17 February 2006.
- [8] Rump, N. (2003): Digital rights management: Technological aspects. In: Becker et al. (2003).
- [9] Pakman, D. (2005): Why DRM everything? A sensible approach to satisfying customers and selling more music in the digital age; Groklaw, 31 December 2005.

Authors' Information

Maria Nickolova – National Laboratory of Computer Virology, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: maria@nlcv.bas.bg.

Eugene Nickolov – National Laboratory of Computer Virology, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: eugene@nlcv.bas.bg.

THE STRATEGICAL IMPACT OF KNOWLEDGE MARKETS AND KNOWLEDGE MANAGEMENT FOR SMALL AND MEDIUM ENTERPRISES

Todorka Kovacheva

Abstract: *According to the rapidly changing environment small and medium enterprises constantly need to adapt their strategies and activities. The transition from the industrial economy to knowledge-based economy results in the increasing of the volume of the available information. Therefore knowledge markets are needed and innovation centers have to be developed. An effective knowledge management system helps small and medium enterprises to overcome their disadvantages and compete with big corporations. The review of current developments in the field of knowledge markets is also made.*

Keywords: *knowledge, knowledge markets, knowledge management system, knowledge-based economy, information society*

Introduction

The globalization and technological developments increase dramatically the information available for the human society. The need of usage of automated tools for the extraction and processing of information appears. This information can be processed through Internet and characterizes with large variety and is unstructured. Therefore the knowledge contained in it can not be extracted and used for the increasing of the effectiveness of the functioning of the contemporary enterprises.

In the continuous and rapidly changing environment with high level of competition for the small and medium enterprises is more and more difficult to survive and growth. They do not have the large amount of financial funds like big corporations. By the usage of the knowledge hidden in the large volumes of freely available information these enterprises can receive high competitive power and successfully compete with big corporations in the virtual and real world market.

The small and medium enterprises can easily become transnational companies which operate on the global market. This is possible as a result of business virtualization, information technologies and telecommunications, Internet and etc.

To use the advantages from the changes in the global economy, small and medium enterprises have two possibilities:

1. to buy knowledge from the knowledge markets, or
2. become a knowledge centers and develop innovations.

Knowledge Markets

According to the fast development of Internet and Information and Telecommunication Technologies causes the appearance of new economy, called Internet or Knowledge Based Economy. A knowledge economy focused on the economy of the producing and management of knowledge. It is also the use of knowledge to produce economic benefits [1]. The term was suggested by Peter Drucker [2]. The Knowledge Economy is not currently absolutely established. Our society is in transition to an information society. And the industrial economy of the 20th century turns into the knowledge-based economy. It is changing the rules of business and competitiveness. The contemporary business competition is based on knowledge. It makes possible for small and starting enterprises to become global and transnational corporations operating with small investments in the world's global market.

The stages in development of the information society are well defined in [3]. The authors in [3] define the difference between information society and other levels of the human been growth by the domination of the

information interests above all others. In the information society the payable information exchange and services dominate above all other market activities. As a result, the Information Market dominates over all other type of markets of the information society. The authors in [3] define the Knowledge Market as a special kind of Information Markets. As the other markets the Knowledge Market is the organized aggregate of participants, which operates in the environment of common rules and principles [4]. The interactions between members of the knowledge markets are described in [5].

A real functioning global knowledge market currently does not exist. There are few attempts to develop such a market but the results are still far from the real knowledge market. Based on the analysis and review of the available scientific literature and information sources the following developments in this field are discovered:

- The theoretical basis of knowledge markets is established in [3, 4, 5, 6, 7, 8, 9, 19, 11, 12, 13, 14, 15]
- An attempt of establishing knowledge markets is made by BRINT [16, 17], The Knowledge Creating Company. It is currently the leading institution in the knowledge market. It is the world's virtual gateway to the largest pool of talented minds involved in contemporary business technology management and knowledge management research and practices across the USA and all other countries of the world.
- Other developments:
 - o The World Bank [18] – Knowledge for Development Program. The Program provides policy advice on four Knowledge Economy pillars: economic and institutional regime, education, innovation, and Information and Communication Technologies (ICTs). It helps countries identify the challenges and opportunities they face in making the transition to the knowledge-based economy. The KAM consists of 81 structural and qualitative variables for 132 countries to measure their performance on the four Knowledge Economy (KE) pillars: Economic Incentive and Institutional Regime, Education, Innovation, and Information and Communication Technologies. Variables are normalized on a scale of zero to ten relative to other countries in the comparison group. The KAM also allows to derive country's overall **Knowledge Economy Index (KEI) and Knowledge Index (KI)**.
 - o The Kaiteur Institute For Knowledge Management [19] develops an e-Knowledge Markets Meta Portal [20].
 - o In April 2006 the Work Foundation [21] announced the start of a major three-year, £1.5 million research programme into the nature of the knowledge economy - the biggest investigation of its kind anywhere in the world. The project aims to identify what the knowledge economy is, and how advanced nations can use knowledge and information to spearhead economic growth and competitiveness in the 21st century.

Knowledge Centers and Knowledge Management System

The transition from the industrial society to the information based society and the transition from the industrial economy to the knowledge-based one logically results in changes in the global business. New industries, new business relationships and new kinds of jobs will appear. To survive in such an environment small and medium enterprises need to have high adaptive capability. Such adaptivity they can receive from the knowledge about the trends and correct forecasts about the changes in the global economical, political and social development. Such knowledge can be established by effective knowledge management system.

Knowing the global business trends small and medium enterprises can become knowledge and innovation centers. Developing innovative products and services they will have a strong competitive position on the global market.

Innovations are deeply connected with the strategical management of the enterprises. Effective strategic management is based on knowledge and correct business predictions. Therefore a knowledge management system is needed. It handles with huge volumes of information collected from various sources and extracts the

hidden knowledge. Knowledge management system combines the latest achievements in the field of information technologies and has a strategic impact over the business development.

The effective knowledge management system needs to have the following functionalities:

- A tool for extracting and processing the information from various internet sources. It needs to be realized as a multiagent system which unites the activities of big number of intelligent robots. They search the Internet space and extract and structure the available information according to specific rules.
- Data bases and knowledge bases.
- Tools for data analysis and knowledge extraction.
- Forecasting tools.
- Querying capabilities, including adhoc queries.

Conclusion

The small and medium enterprises can be competitive with big corporations by using an effective knowledge management system. Such a system helps them to overcome the missing financial funds needed for their adaptation in the constantly changing environment. Becoming a knowledge and innovation centers small and medium enterprises can dramatically increase their competitive power and their impact on the global market.

Bibliography

- [1] http://en.wikipedia.org/wiki/Knowledge_economy
- [2] Peter Drucker, (1969). The Age of Discontinuity; Guidelines to Our changing Society. Harper and Row, New York. ISBN 0-465-08984-4
- [3] Markov K., Ivanova K., Mitov I., The Staple commodities of the knowledge market, International Journal "Information Theories & Applications", Vol.13, Number 1, pp.11-18, 2006
- [4] K. Markov, K. Ivanova, I. Mitov, N. Ivanova, A. Danilov, K. Boikatchev. Basic Structure of the Knowledge Market. IJ ITA, 2002, V.9, No.4, pp.123-134 (Presented at Int. Conf. ICT&P, 2002, Primorsko)
- [5] Ivanova K., Ivanova N., Danilov A., Mitov I., Markov K., Basic interactions between members of the knowledge market, International Journal "Information Theories & Applications" Vol.13, Number 1, pp. 19- 30, 2006
- [6] N. Ivanova, K. Ivanova, K. Markov, A. Danilov, K. Boikatchev. The Open Education Environment on the Threshold of the Global Information Society. IJ ITA, 2001, V.8, No.1 pp.3-12. (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.272-280, in Russian, Presented at Int. Conf. ICT&P 2001, Sofia, pp.193-203)
- [7] Кр.Иванова, Н.Иванова, А.Данилов, И.Митов, Кр.Марков. Обучение взрослых на рынке профессиональных знаний. Сборник доклады на Национална научна конференция "Информационни изследвания, приложения и обучение" (i.TECH-2003), Варна, България, 2003. Стр. 35-41.
- [8] K.Markov. A Multi-domain Access Method. Proceedings of the International Conference on Computer Based Scientific Research. PLOVDIV, 1984, pp.558-563.
- [9] Кр.Марков. От миналото към бъдещето на определението на понятието "информация". Сборник "Програмиране '88". БАН, Варна, Дружба, 1988, стр.150.
- [10] K.Markov, K.Ivanova, I.Mitov. Basic Concepts of a General Information Theory. IJ ITA {International Journal "Information Theories & Applications"}; 1993; v.1, n.10; pp.3-10.
- [11] Kr. Markov, Kr. Ivanova, I. Mitov, N. Ivanova, K. Bojkachev, A. Danilov. Co-operative Distance and Long-Live Learning. ITA-2000, Bulgaria, Varna, 2000.
- [12] K. Markov, P. Mateev, K. Ivanova, I. Mitov, S. Poryazov. The Information Model. IJ ITA, 2001, V.8, No.2 pp.59-69 (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.465-472)
- [13] K. Markov, K. Ivanova, I. Mitov, N. Ivanova, A. Danilov, K. Boikatchev. Basic Structure of the Knowledge Market. IJ ITA, 2002, V.9, No.4, pp.123-134 (Presented at Int. Conf. ICT&P, 2002, Приморско)
- [14] Kr.Markov, Kr.Ivanova, I.Mitov. General Information Theory. Basic Formulations. FOI-COMMERCE, Sofia, 2003. ISBN 954-16-0024-1.
- [15] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. Formal Definition of the Concept "INFOS". Int.Journal "Information Theories and Applications", 2004, ISSN 1310-0513. V.11, N.1, pp.16-19. (Presented at i.TECH2004, Varna, Bulgaria. pp. 71-74)
- [16] <http://www.brint.com/wwwboard/wwwboard.html>
- [17] <http://www.brint.org/>

[18] <http://web.worldbank.org/WBSITE/EXTERNAL/WBI/WBIPROGRAMS/KFDLP/0,,menuPK:461238~pagePK:64156143~piPK:64154155~theSitePK:461198,00.html>

[19] <http://www.kikm.org/portal/page2.htm>

[20] Featured in IEEE Computer Society ITProfessional Magazine March-April 2001 Issue

[21] <http://www.theworkfoundation.com/futureofwork/research/knowledgeeconomy.aspx>

Author's Information

Todorka Kovacheva – Gluon Technologies Ltd, Varna, Bulgaria, www.gluontechnologies.com,
e-mail: todorka_kovacheva@yahoo.com, phone: +359899920659

AN APPROACH TO REPRESENTING THE PROCESS OF INFORMATION BUSINESS MODELING

Nadezhda Filipova, Filcho Filipov

Abstract: *The compact and visualized documenting of information business modeling is a major prerequisite for comprehending its basic concepts, as well as for its effective application and improvement. The documenting of this process is related to its modeling. Thus, the process of information business modeling can be represented by its own tools. Being based on this thesis, the authors suggest an approach to representing the process of information business modeling. A profile for its documenting has been developed for the purpose.*

Keywords: *business modeling, SPEM, UML profile*

ACM Classification Keywords: *I.6.5 Model Development - Modeling methodologies*

Introduction

Information business modeling (iBM) aims at visual representation of the business processes of the target organization by means of methods for information system development [Filipova, 2003]. This is a methodology for analysis and reengineering of organizations, as well as for developing adequate integrated computer information systems (CIS). Hence, iBM is a common business process, through which one could represent business processes, performed in organizations, including these for analysis and design, CIS development methodology, etc. This means that iBM is a process for modeling and representing other processes, i.e. this is a metaprocess. This basic feature of iBM emphasizes its importance, and the necessity to understand its components, tools and mechanisms.

According to us, the metamodel of iBM can be the basis to represent the process of its realization. Our researches in the field of the system development metaprocess, and our experience in object-oriented modeling make us assume that the iBM metamodel can be defined, using the SPEM metamodel and the UML profile mechanism.

In this context, our goal is to suggest an approach to representing the iBM process in a compact and visual way. For the purpose, a profile for the process of iBM will be defined on the first place, and then it will be transformed into a profile for its documenting.

I. A framework for representing the process of iBM

iBM is a process that can be represented by the object-oriented approach. According to the modern concepts, a four-layer architecture can be used for object-oriented modeling of real processes and phenomena. Its layers are in hierarchical order, and each one can be represented through the concepts of the upper one (Table 1). We must

note, that this is an architecture for representing both the process and the product or the system, created in its application.

The dotted line in Table 1 shows the object of our research, namely layers M2 and M1 of the four-layer architecture.

The metamodel of the iBM process is a subset of SPEM [SPEM, 2005], on one side, and its specialization – on the other. The product at a metamodel level is represented by means of UML. The product of iBM however is specific, and in order UML to be applied adequately, the latter must also be specialized. The method specialization at the M2 layer may be accomplished through the UML profile¹ mechanism.

Therefore, the metamodel of the process of iBM can be defined as a specialization of the SPEM metamodel summarized in **a profile for the process of iBM**. Likewise, the UML possibilities applied in iBM can be summarized in **a profile for iBM**. Thus the M2 layer of the architecture is divided into two sublayers: a layer of the metamodel (M2.1), and a layer of the profiles (M2.2). M2.1 comprises the metamodels of UML and SPEM, whilst M2.2 includes the profile for iBM and the profile for the process of iBM. These two profiles exactly outline the framework for iBM representation. This framework is used to describe the model of the process of iBM at the M.1 layer of the architecture for object-oriented modeling. Aimed at the more compact and visual representing of iBM, we will use just a part of the profile for the process of its implementation, encapsulated into a profile for its documenting.

In order to document iBM on the basis of the framework defined, it is necessary to explore in details the two profiles, mentioned above, namely the profile for iBM and the profile for the process of iBM.

Table 1. A four-layer architecture for object-oriented modeling

Layer	Name	Contents
M3	Metametamodel of the process	MOF(Meta Object Facility) – integrates methods and processes into a common framework. The metamodels in MOF are represented by a subset of the UML
M2	Metamodel of the process ²	UML, SPEM , basic concepts / metamodel of iBM
M1	Model of the process	A concrete instance of the CIS development methodology – e.g. IBM RUP [Kruchten, 2003; Rational 2003], OPEN [Henderson-Sellers, 2000], MSF (Microsoft Solution Framework) [Duffy, 2003], XP, iBM, etc.
M0	Executable process	A real process for implementing a project

II. Components of the profile for iBM

The profile for iBM is discussed in [Filipova, 2003; Kruchten, 2003; Rational 2003]. Its components are classified into three groups, as follows: model elements, models, diagrams. The information business models are comprised of model elements, and are represented as various diagrams.

The **model elements** include: Business use case () , Business actor () , Scope () , Business goal

¹ The profile is defined as “lightweight extensibility mechanism”, consisting of stereotypes, tagged values, and constraints

² This layer is also called a method layer

() , Business worker () , Business use case realization () , Business entity () , Business system () , Business event () , Business rule () . They are specializations of corresponding UML model elements.

Three major **models** are created in iBM:

- 1) Business use case model (BUCM) – this model reflects the business goals and intended functions of the organization, i.e it answers the question “What is done”. The model is used to define the roles of the organization, and the products delivered. It represents the work of the organization as a set of business use cases, i.e. business processes.
- 2) Business analysis model (BAM) – represents the internal aspects (i.e the realization) of business use cases by modeling the interaction between business workers and business entities.
- 3) Object business model (OBM) – this is a partial BAM, including just business entities, but not the responsibilities of business workers. This model reflects the static aspects of the processes explored.

The static and dynamic aspects of information business models are visualized by UML diagrams, which are used in a specific way. They are summarized in Table 2.

Table 2. Diagrams of iBM at model levels

Model	Diagrams	
	Static	Dynamic
BUCM	- Business use case diagram	- Activity diagram
BAM	- Business class diagram	- Activity diagram - Business sequence diagram - Business communication (collaboration) diagram - State machine diagram
OBM	- Business entity diagram	

III. Defining the profile for the process of iBM

The initial prerequisites for defining the profile for the process of iBM follow:

the first, iBM is a business process;

the second, this profile is a subset of the SPEM profile, and interprets its components in a specific way;

the third, this profile uses as artifacts the components of the profile for iBM.

Our first conclusion, derived from the first prerequisite, is that the process can be represented by means of the profile for iBM. This is not a good decision however, as there will be overlapping between the representation of the process of iBM, on the one hand, and the representation of its product – on the other. Moreover, this can produce confusion in iBM process documenting. The stated problem was confirmed by our preliminary experiments on modeling the iBM process. Thus, it is necessary to define a specialized profile for representing the process of iBM.

Our second conclusion (resulting from the first prerequisite), which is also our thesis, is that there is a direct connection between the profile for iBM and the profile for the process of iBM. Therefore, we should find the correspondence between their components. And the second prerequisite prompts that these components are inherited from the SPEM profile.

We will concern the consequences of the third prerequisite when discussing the profile for iBM documenting.

The initial prerequisites for defining the profile for the iBM process direct our strategy, i.e. we are going to seek semantic equivalence between the components of the iBM profile, on the one hand, and these of the SPEM profile, on the other. Furthermore, this comparison will be accomplished at levels of abstraction, i.e. models.

The major components of the SPEM profile are: Goal () , Workflow / Workflow details () , Phase / Iteration () , Artifact () , Process package / Discipline () , Process / Life cycle () , Role () , Activity () .

Following our strategy for defining the profile for the process of iBM, we must achieve complete equivalence between the components of the profiles, that we compare, for the first of the models, namely the Business use case model. We establish that there is neither Scope, nor Business actor in the SPEM profile. Our answer to this problem is:

- first, introduce a Scope component in the profile for the process of iBM with its meaning and icon, inherited from the iBM profile;
- second, in order to introduce a Business actor component however it is necessary to analyze its semantic. The business actor is a *user* of the products of the process of iBM. Besides, he is an external participant in this process, assisting in its implementation. Therefore, the Business actor is a user of the process, and a kind of a role with limited responsibilities. That's why we introduce a new stereotype, named Process user

() , in the profile for the iBM process.

Unlike the Business use case model, we will seek just partial equivalence of the components of the Business analysis model and the Object business model. We find out difference in several components at BAM level, and to be precise these are: Business use case realization, Business event, Business rule, Activity.

The Business use case realization is a collaboration¹, i.e. it groups a set of dynamic and static diagrams, reflecting structure and behavior of a business use case. In this case the collaboration shows how a certain elementary process (a subprocess) of the iBM process is implemented through the interaction of activities, roles and artifacts, i.e. the workflow details are described. That's why we introduce a new stereotype in the profile for the process of iBM, which is similar to the Business use case realization -

Workflow realization () .

The documenting experience, gathered in some methodologies, e.g. IBM RUP, proves that components such as Business event and Business rule are rarely used. Hence we will not look for their equivalences, and they will not be used in the profile for the iBM process.

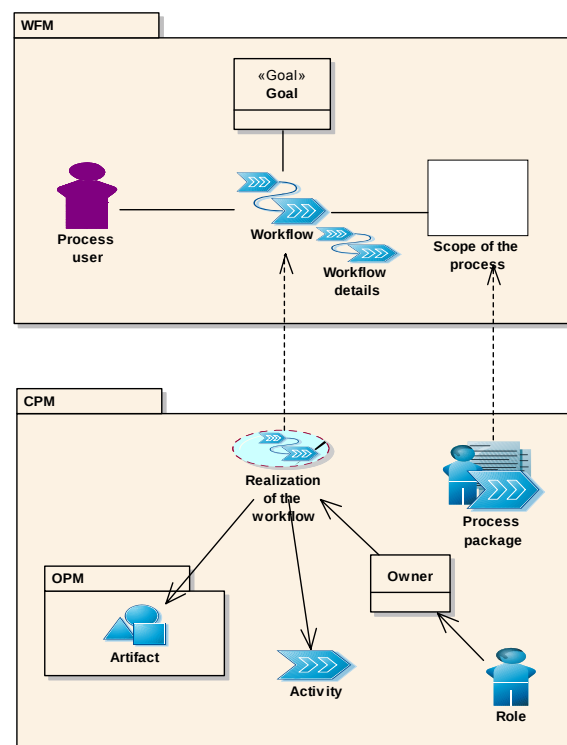


Fig. 1. Models and components of the profile for the process of iBM

¹ this is a standard UML component to implement behavior

We must point that the Activity component of the profile for the process of iBM is an operation of the Business worker in the iBM profile.

The models in the profile for the process of iBM acquire different manifestations, which is a result of the semantic of their components. Thus, the workflow is a basic component of the Business use case model, which makes us name it Workflow model (WFM). Its static aspects are represented by Workflow diagram (analogous to Business use case diagram – Table 2), and its dynamic aspects are represented by Activity diagram. Concerning similar considerations, the Business analysis model is named Conceptual process model (CPM), and the Object business model – Object process model (OPM).

After these comments and elaborations, we can represent the components of the profile for the iBM process, and the relations among them (Fig.1).

IV. Transforming the profile for the process of iBM into a profile for its documenting

Regarding the concepts of iBM, we assume it is adequate to build up a Business use case model, i.e. a Workflow model, in order to present a process which is both highly abstract and of wide scope, on the one side, and purposeful and easy to be understood, on the other. Hence, the profile for documenting the iBM process must encompass all the components of the WFM.

According to us, it is necessary to use the product model, represented by the iBM profile, in order to achieve greater purposefulness when modeling the workflows in iBM. This means that the components of the iBM profile are artifacts of the iBM process, and that only a part of the Conceptual process model will be used. The roles and activities of the iBM process will be used unstructurally, i.e. the relations among them are not going to be represented.

The profile for documenting the iBM process includes also the models and diagrams used. They are encapsulated into a package, named iBM models and diagrams.

The profile for documenting, we have defined, is depicted on fig.2. We must point again that the Conceptual process model of iBM has a wider

scope. The profile for documenting however includes just a part of the Object process model, namely the packages Profile for iBM and iBM models and diagrams, and the Role and Activity components.

Using the profile for documenting defined, we have made some experiments to model the process of iBM, which are based on the Business modeling workflow of IBM RUP. To be exact, we have developed a context diagram of iBM, a mechanism for representing the decomposition of processes into subprocesses, based on decomposition diagrams, a template for specifying the iBM subprocesses (fig. 3). On this basis the iBM subprocesses have been documented, Workflow diagram and Activity diagram have also been built up.

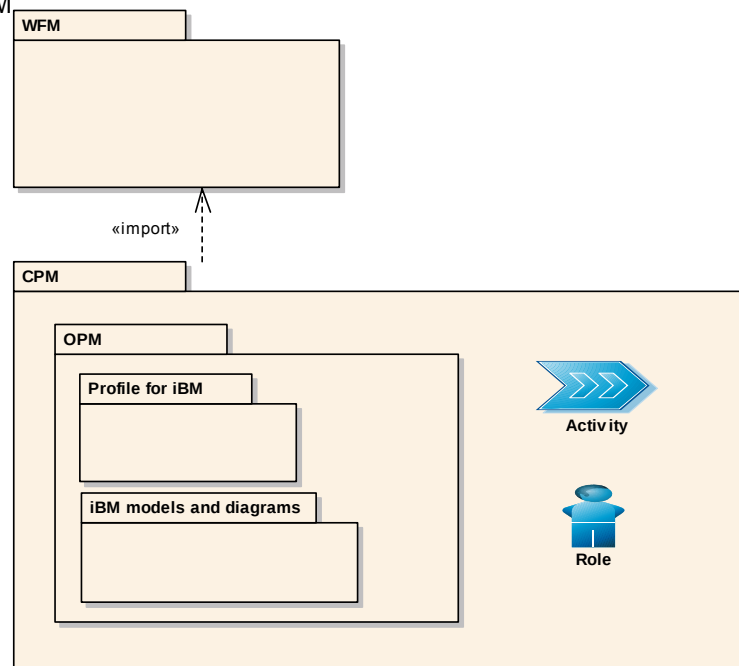


Fig.2. Profile for the documenting the process of iBM

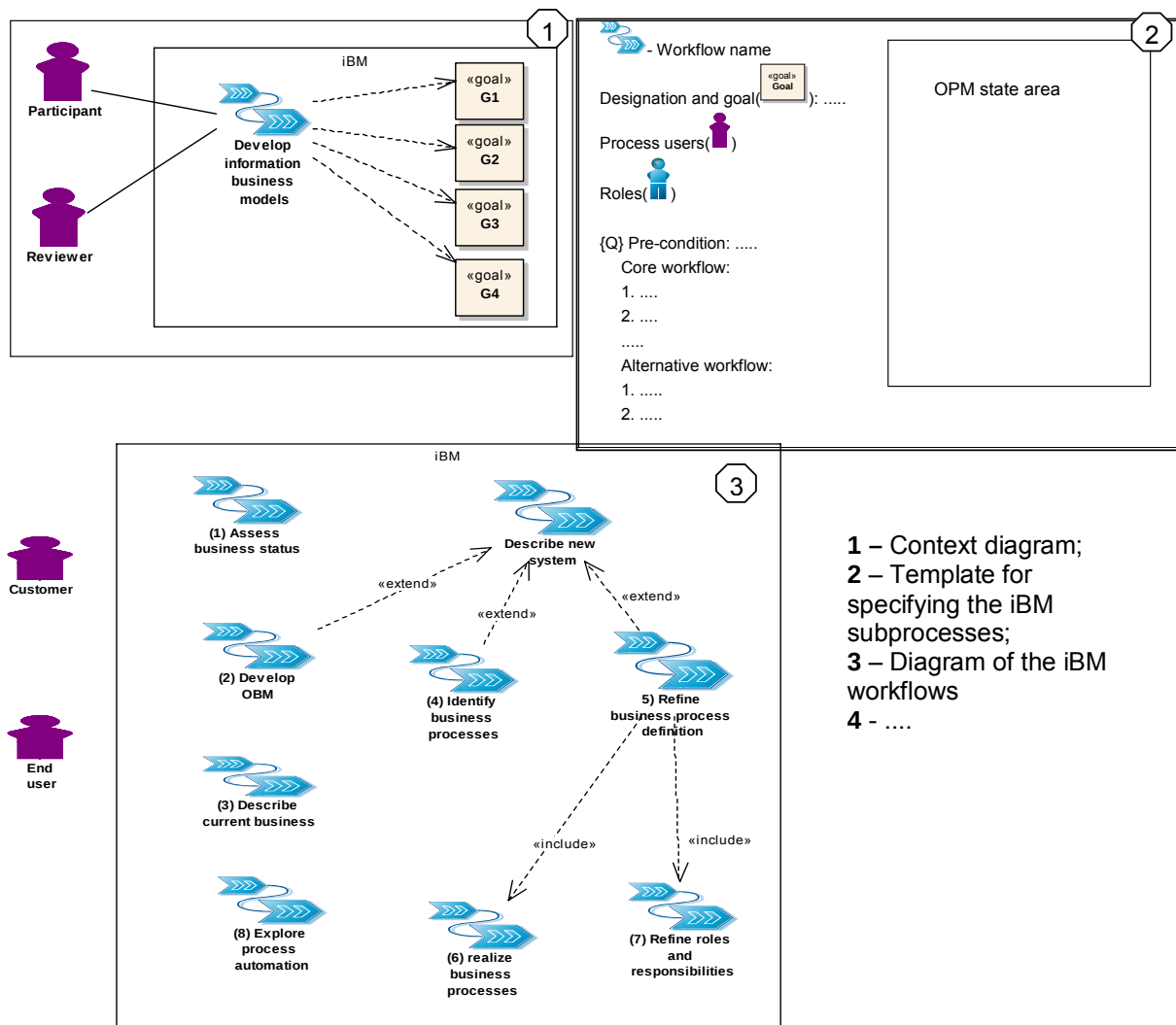


Fig. 3. Diagrams and templates in documenting the iBM process

Conclusion

We will summarize the steps fulfilled in defining the profile for documenting the iBM process. On the first place, we have identified the components of the framework for representing the iBM process, i.e. the profile for iBM, and the profile for the process of iBM. Then the profile for the process of iBM has been defined, which was an important target of ours. This profile emerged on the basis of the SPEM profile, and was driven by our thesis for semantic correspondence between the profile for the product and the profile for the process of iBM. Afterwards, the profile for the iBM process has been transformed into a profile for its documenting, regarding the requirements to the model of the iBM process. Using this profile, we have made some experiments to represent the process of iBM. More precisely, we have built up the iBM workflow model and its subprocesses have been specified by the template defined for the purpose.

Bibliography

- [Duffy, 2003] Duffy, Scott, David Waddleton. MCSA Analyzing Requirements and Defining .NET Solution Architectures Study Guide (Exam 70-300). McGraw-Hill/Osborne, 2003
- [Filipova, 2001] Filipova, N., F. Filipov. Documenting Object-Oriented Profiles. Int. Journal of Information Theories & Applications, Vol.7, No.4, pp 175-182

- [Filipova, 2003] Filipova, N. Basic concepts of information business modeling (in Bulgarian). Известия /Списание на ИУ-Варна/, бр.4, 2003 г. , с. 27
- [Filipova, 2005] Filipova, N. Possibilities of the UML for information business modeling (in Bulgarian). Годишник на Икономически Университет - Варна, том 77, 2005 г.
- [Henderson-Sellers, 2000] Henderson-Sellers, B. The OPEN Framework for Enhancing Productivity. IEEE Software, March/April, 2000
- [Kruchten, 2003] Kruchten P. The Rational unified process – an introduction. Addison Wesley, Third Ed., 2003
- [Rational, 2003] Rational Unified Process. Version 2003, Rational Software Corporation, 2003
- [SPEM, 2005] Software Process Engineering Metamodel Specification. Version 1.1, OMG Group, 2005
-

Authors' Information

Assoc. prof. Nadezhda Filipova - University of Economics – Varna, Bul. Kniaz Boris I # 77, Varna 9003; e-mail: Filipova@ue-varna.bg

Assoc. prof. Filcho Filipov - University of Economics – Varna, Bul. Kniaz Boris I # 77, Varna 9003; e-mail: FFilipov@ue-varna.bg

THE ROLE OF DBMS IN ANALYTICAL PROCESSES OF THE LOGISTIC OF STOCK RESERVES

Julian Vasilev

Abstract: *One of main problems of corporate information systems is the precise evaluation of speed of transactions and the speed of making reports. The core of the problem is based on the DBMS that is used. Most DBMS which are oriented for high performance and reliability of transactions do not give fast access to analytical and summarized data and vice versa. It is quite difficult to estimate which class of database to use. The author of the article gives a concise observation of the problem and a possible way to be solved.*

Keywords: Database management systems (DBMS), Information technology, Cache, Interbase.

ACM Classification Keywords: H.2.8 Database Applications, H.4 information systems applications.

Introduction

Stock reserves affect all activities in the enterprise. Their management is directly connected to the appliance of different methods. According to Gatorna [1,333] "...logistics includes two main spheres of activity: management of materials entering the production and the management of distribution of final products."

According to the information needs of the operational management we meet requirement of different groups of users who give priorities for fastness of defined actions. For instance, people who get stock from suppliers and store it in the factory warehouse want high speed of their transactions. By analogy with supplies, the sales department needs high speed in making, saving and printing invoices. Having in mind these requirements, the key role of the DBMS (Database management system) is to provide high speed of transactions. As we mentioned these transactions concern registering documents in the information system of the enterprise. In practice certain raw materials are stored and "wait" their participation in production. At the end of the enterprise final products form also stock reserves. This type of production activities is usual for Bulgarian enterprises. Without involving logistics, these activities are organized in accounting software.

Layout

According to managers, production operations are subject to monitoring using certain indicators, such as effectiveness, profit, costs, revenues. A full control of material flows is a precondition for the science logistics to give a solution to several problems in the sphere of creating an order for supply. Supply department needs to know the amount of order, the frequency of sending, stocks included, supplier. Analytical processes refer to a higher level of data aggregation and extracting synthesized data. Some authors [2,100] give account of necessity of expert systems in logistics for solving complex problems. According to other experts [3,182] in building “computers are used for solving complex problems of planning and for techno-economical problems when building plans for material and technical providence”. According to other authors [5] “multi-dimensional data structures are the base of the conception of direct analytical process”. Top managers are interested in the dynamics of several indicators in order to monitor the state of the organization. The warehouse of the enterprise generates big amounts of data. Data increases throughout the time. Observations in practice show that new technical data storage devices are with bigger volume, than their precedents, so data storage is not a problem for IT specialists. Problems occur when we need to show dynamic indicators. The use of DBMS, accenting to high speed of transactional performance, for example Cache (a product of Intersystems), guarantees quick and reliable registration of data. DBMS of that class usually make reports slower than other classes of DBMS such as Oracle or MS SQL Server. The application of last two databases is combined with a spread application development instruments for OLAP (on-line analytical data processing) data analysis, which make them adjusted to solve managers’ problems of high level. Information systems based on Oracle require significant hardware resources. Another disadvantage is the speed of transactions. Keeping indexed data in relational structures is a requirement for fast speed of OLAP instruments. But DBMS need much time to keep indexed data actual. With increase of data more time is needed for a transaction to be saved. We made an observation. Results are summarized in Table 1.

Table 1: Comparison between Cache, MS SQL Server and MS Access

DBMS/Indictor	Cache	MS SQL Server	MS Access
SQL access	yes	yes	yes
Support of huge data structures	yes	yes	yes
Speed of grouping and summarizing data	low	high	middle
Support of OLAP	no	yes	no
Speed of saving transactions	high	middle	low

We consider that the key problem of the DBMS role is the seeking and finding of an objective compromise between high speed of transactions and high speed of analytical processes in management. Most IT experts prefer popular DBMS such as Oracle, MS Access, Interbase, DB2, MS SQL Server. Advertisements in IT magazines and newspapers, application development environments make them attractive. Managers need to monitor a set of economic indicators concerning activities in the enterprise. The compromise is oriented to high speed of registration accounting documents from one side and to the logistics on the other side. A huge document flow requires high speed of transactional performance. If it is not provided, customer services are slowed down. Moreover the company needs personnel to process these documents. Customer services have to be fast and with high quality. These are main priorities of marketing. Whereas future sales are subject to research of the science “Forecasting”, future supplies are a matter of organization of the procurement department and are subject of discussion by scientists in the sphere of logistics. Forecasting sales and organization of future supplies is deeply connected to analysis of big arrays of data, meaningless methods which can be applied. Backing our opinion other authors [4, 57] think that “Forecasting of material requirements is based on values from historical data. Proving of future needs is helped by statistical methods”

For the purpose of our research we made a comparison between Cache and Interbase. Results are summarized in Table 2.

Table 2: Comparison between Cache and Interbase

DBMS/Indicator	Cache	Interbase
Time to insert 100 000 records in the database, each is 255 bytes long	8 seconds	350 seconds
Filtering 100 000 records and retrieving 5 000 records by a given criteria	4 seconds	8 seconds

It is indisputable that there is a wide choice of DBMS. In accordance with the opinion of some authors [6] "For building databases commercial DBMS can be used such as ORACLE, MS-SQL, DB2 and others, as well as free versions." We consider that for building corporate systems IT experts can not rely on DBMS which does not have support. Experiments were done on Win XP, with 128 MB RAM, Intel Celeron processor 466 MHz. To visualize resulting data we used following technologies: HTML (hyper text markup language), ASP (active server pages) and GUI (graphic user interface) application. The GUI application visualizes data faster than ASP technology which needs an IIS (Internet information server).

Conclusions

Estimating DBMS we need to bare in mind not only its popularity but its orientation to fast transactions or high speed of analytical processes. The right choice is based on logistic processes, hardware resources, personnel and an objective forecast for the increase of data arrays. The last factor is usually ignored and sometimes it is decisive. Its correct evaluation is done after several years of functioning of a store or accounting system. One of the ways to solve this problem is to be built an application server, which acts as a Windows process which transports data from one DMBS to another. For instance we can have an information system based on Cache, and a second one based on MS SQL Server. The first one is used for registering documents such as invoices and stock receipts, the second one for OLAP analysis. The mediator is the upper mentioned application server. In our opinion it is a temporary solution until further development of DBMSs.

Bibliography

- [1] М Гаторна, Д. Основи на логистиката и дистрибуцията, Бургас, 1996.
- [2] Гаджинский, А. Логистика, Москва, 2000.
- [3] Генов, Г. Логистика в строителството, ИУ Варна, 1998.
- [4] Коралиев, Я. Логистика на снабдяването, София, 1998.
- [5] Несторов, К. Организация на данни в програмни продукти за счетоводно отчитане, сп. Известия, ИУ Варна, бр. 2, 2002г.
- [6] Петров, П. Проблеми на изграждането на автоматизирани информационни системи в големи производствени предприятия сп. Известия, ИУ Варна, бр. 1, 2002г.

Author's Information

Julian Vasilev – Chief assistant professor, Department of Informatics, Varna University of Economics; 77, Kniaz Boris I str.; Varna; Bulgaria; e-mail: vasilev@ue-varna.bg

BIVIRTUAL ORGANIZATION AS A QUEUING SYSTEM

Tetiana Palonna, Iurii Palonnyi

Abstract: The main features of virtual organizations are outlined. The mathematical models of functioning of virtual organization are offered on the basis of theory of queuing systems. Characteristics of efficiency are examined.

Keywords: virtual enterprise, virtual laboratory, queuing system.

ACM Classification Keywords: I.6.4 Model Validation and Analysis; K.6.1 Project and People Management - Systems analysis and design.

Introduction

Saturation of goods markets and development of information technologies made possible such kind of organization of productive activity as virtual organization.

In the present work we shall consider such subspecies of virtual organization as a virtual laboratory.

There have been suggested mathematical models of virtual organization functioning on the basis of the queuing theory. The efficiency characteristics are considered.

Problem Statement

Let us consider virtual organization as the combination of three interconnected components: local agents (LA), upper agents (UA) and service [Dawidow, 1992]. For virtual laboratory these components could be interpreted in the following way: local agents – students, upper agents – the educational institution and a server.

In case of a server work we assume, that the assembly of agents represents the source of queries of unlimited capacity, and a server is a queuing system (QS), assigning time for the queries processing [Minzberg, 2001].

The average time for one request handling t_{serv} is composed of the average time of the start of connection t_{start} , time of connection $\bar{t}_{connect}$ and the time of the end of connection $\bar{t}_{end}$: $\bar{t}_{serv} = \bar{t}_{start} + \bar{t}_{connect} + \bar{t}_{end}$, at that the meanings of $\bar{t}_{start}$ and $\bar{t}_{end}$ are small comparing with the value of $\bar{t}_{connect}$. The intensity of handling is a value, inverse to the average time of handling, so we receive:

$$\mu = \frac{1}{\bar{t}_{serv}} = \frac{1}{\bar{t}_{start} + \bar{t}_{connect} + \bar{t}_{end}}.$$

The intensity of arrival of queries from agents we shall indicate as λ .

For the sake of simplification we assume, that the server is the one-channel QS with the refusals. Let's assume that there exist two standard situations in the server operation: 1) the "hot season", when the server can not handle with the flow of requests (such faults could, for example, arise before the session, when the large quantity of lower agents apply with the queries); 2) the "dead season" or the vacations, when the intensity of arrival of service requests sharply drops.

The state of the QS being considered is determined after the number of queries in it:

- 1) for the virtual organization:
 - S_0 - in the system there is no queries;
 - S_1 - in the system there is one query and the UA is handling it.
- 2) for the virtual laboratory:

S_0 – in the QS there is no queries (the server stands idle);

S_1 – in the QS there is one query (the server is busy with the handling of the given query).

The transition from S_0 to S_1 takes place depending of the intensity of requests and the probability of the connection confirmation $p_{connect}$.

From the given description of the QS functioning we receive, that the density $d_{01}(t)$ of the system transition from the state S_0 into the state S_1 is equal to the product of $p_{connect}$ by the intensity λ of the incoming flow of queries, and the density $d_{10}(t)$ of the transition from the state S_1 into the state S_0 – to the intensity μ of the flow of queries handling. That is why the transition graph will be of the form, represented at Fig. 1

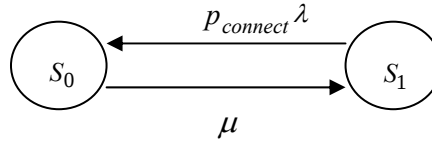


Fig. 1. Transition graph of the interaction model of agents and the server (for virtual laboratory)

Let us assume that all the flows of events in the QS are the simplest. Than in the QS the Markovian process takes place. After the transition graph we receive, that the functioning of the QS is depicted by the Chapman-Kolmogorov differential equation system [Kremer, 2001]:

$$\begin{aligned}\frac{dp_0(t)}{dt} &= -p_{connect}\lambda p_0(t) + \mu p_1(t), \\ \frac{dp_1(t)}{dt} &= -\mu p_1(t) + p_{connect}\lambda p_0(t),\end{aligned}$$

together with the normalization requirement

$$p_0(t) + p_1(t) = 1.$$

We consider, that at the initial moment of time in the system there were no queries, i.e. the system was in the state S_0 :

$$p_0(0) = 1, p_1(0) = 0.$$

Excluding the second equation of the system and using the normalization requirement we receive the ordinary differential equation:

$$\frac{dp_0(t)}{dt} = -(p_{connect}\lambda + \mu)p_0(t) + \mu.$$

Analytical solution of the mentioned equation taking into account the initial conditions has the looks in the following way:

$$p_0(t) = \frac{\mu}{p_{connect}\lambda + \mu} + \frac{p_{connect}\lambda}{p_{connect}\lambda + \mu} e^{-(p_{connect}\lambda + \mu)t}, p_1(t) = 1 - p_0(t).$$

Similarly after the transition graph we can obtain the balance equation (for the sake of establishing the working regime of the system):

$$p_{connect}\lambda p_0 = \mu p_1, \mu p_1 = p_{connect}\lambda p_0, p_0 + p_1 = 1,$$

from where the values of probabilities of the system states are calculated:

$$p_0(t) = \frac{\mu}{p_{connect}\lambda + \mu}, p_1(t) = \frac{p_{connect}\lambda}{p_{connect}\lambda + \mu}.$$

Let us consider the simplest model of interaction of the LA and the service. If ν is the intensity of offers from service to the LA, and η is the intensity of theirs accepting by the LA, then there are two stable states of the system: S_0^1 - the proposal is accepted, S_1^1 - the proposal is not accepted.

Model of interaction of the agent and the server is considered similarly. Let ν be the billing intensity by the server to the agent; and η – the intensity of their payment by the agent. As earlier, let us assume, that all the flows of the events are the simplest flows and consequently in the system the Markovian process takes place. Two stable states if the system considered could be distinguished: S_0^1 – the bill is paid, S_1^1 – the bill is unpaid.

Then the transition graph will be of the form, represented at Fig. 2.

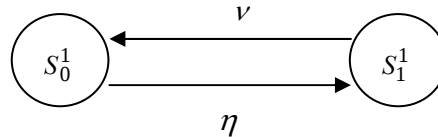


Fig. 2. Transition graph of the interaction model of the local agents and the service

Solving the given problem similarly to the previous one, we shall receive the probability values of the QS state for the steady-state working regime

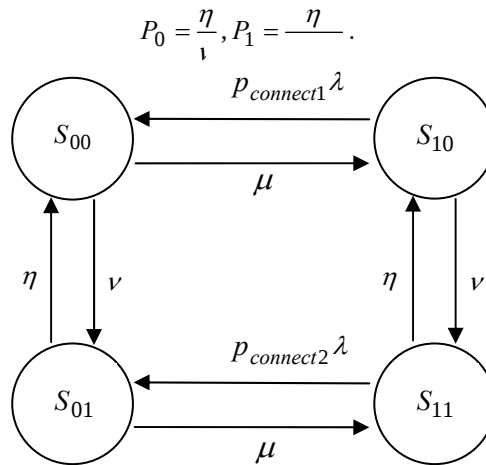


Fig. 3. Transition graph of the integrated model (of the fragment of virtual organization)

Integrating the received simplest modes we shall obtain the QS with four states:

- 1) for virtual organization:
 - S_{00} - the UA is free, the proposal has been accepted;
 - S_{01} - the UA is free, the proposal has not been accepted;
 - S_{10} - the UA is occupied, the proposal has been accepted;
 - S_{11} - the UA is occupied, the proposal has not been accepted.
- 2) for virtual laboratory:
 - S_{00} - the server is free, the bill is paid;
 - S_{01} - the server is free, the bill is unpaid;
 - S_{10} - the server is occupied, the bill is paid;
 - S_{11} - the server is occupied, the bill is unpaid.

Describing functioning of the QS as changes of its states we can obtain the transition graph, represented at Fig. 3.

Balance equation

After the transition graph we obtain the mathematical model of the system functioning - the balance equation:

$$\begin{aligned}(p_{connect1}\lambda + \nu)p_{00} &= \mu p_{10} + \eta p_{01}, \\ (\mu + \nu)p_{10} &= p_{connect1}\lambda p_{00} + \eta p_{11}, \\ (\eta + p_{connect1}\lambda)p_{01} &= \nu p_{00} + \mu p_{11}, \\ (\mu + \eta)p_{11} &= p_{connect2}\lambda p_{01} + \nu p_{10}, \\ p_{00} + p_{10} + p_{01} + p_{11} &= 1.\end{aligned}$$

Solution of the given system of linear algebraic equations looks in the following way:

$$\begin{aligned}p_{00} &= \frac{1 - \alpha_2 p_{11}}{\alpha_1}, \\ p_{01} &= \frac{1 - \beta_2 p_{11}}{\beta_1}, \\ p_{10} &= \frac{1 - \gamma_2 p_{11}}{\gamma_1}, \\ p_{11} &= \frac{2}{\left(\frac{\alpha_2}{\alpha_1} + \frac{\beta_2}{\beta_1} + \frac{\gamma_2}{\gamma_1}\right)},\end{aligned}$$

where

$$\begin{aligned}\alpha_1 &= 1 + \frac{\nu}{\eta + p_{connect2}\lambda} + \frac{p_{connect1}\lambda}{\mu + \nu}, \\ \alpha_2 &= 1 + \frac{\mu}{\eta + p_{connect2}\lambda} + \frac{\eta}{\mu + \nu}, \\ \beta_1 &= 1 + \frac{\eta + p_{connect2}\lambda}{\nu} + \frac{p_{connect2}\lambda}{\nu}, \\ \beta_2 &= 1 - \frac{\mu}{\nu} + \frac{\mu + \eta}{\nu} = 1 + \frac{\eta}{\nu}, \\ \gamma_1 &= 1 + \frac{\mu + \nu}{p_{connect1}\lambda} - \frac{\nu}{p_{connect1}\lambda} = 1 + \frac{\mu}{p_{connect1}\lambda}, \\ \gamma_2 &= 1 - \frac{\eta}{p_{connect1}\lambda} - \frac{\mu + \eta}{p_{connect2}\lambda} = 1 - \frac{\mu}{p_{connect2}\lambda}.\end{aligned}$$

Characteristics of functioning of virtual laboratory (virtual organization)

After the proposed model the following system characteristics are calculated:

- probability of the refusal because of the bill, unpaid by the agent
 $P_{ref,unb} = p_{00}(1 - p_{connect1}) + p_{01}(1 - p_{connect2})$
- probability of the refusal because of the occupancy of the server (consequently, the UA)
 $P_{ref,s} = p_{10} + p_{11}$
- probability of the refusal of the servicing $P_{ref} = P_{ref,unb} + P_{ref,s}$
- relative throughput $q = 1 - P_{ref} = p_{00}p_{connect1} + p_{01}p_{connect2}$
- absolute throughput $A = \lambda q$
- average time of staying of the query in the system

$$\bar{Z} = \frac{A}{\mu} = \frac{\lambda}{\mu} q = \rho(1 - P_{ref}) = \rho p_{00} p_{connect1} + \rho p_{01} p_{connect2}$$

Conclusion

The mathematical models of functioning of virtual organizations, namely of virtual laboratories in the capacity of one-channel queuing system with the refusals were considered. The characteristics of functioning of the considered objects were analysed.

Bibliography

- [Kremer, 2001] Н.Ш. Кремер, Б.А. Путко, И.М. Тришин. Исследование операций в экономике. Москва, ЮНИТИ, 2001.
- [Dawidow, 1992] Dawidow W., Malone M. The Virtual Corporation. Structuring and Revitalizing the Corporation for the 21st Century. – New York, Harper Collins, 1992
- [Minzberg, 2001] Г. Минцберг. Структура в кулаке. Создание эффективной организации. Санкт-Петербург, 2001.

Authors' Information

Tetiana Palonna - Cherkasy State Technological University, Shevchenko blv., bl.460, Cherkasy-18006, Ukraine;
e-mail: tansha@ukr.net

Iurii Palonnyi - Cherkasy State Technological University, Shevchenko blv., bl.460, Cherkasy-18006, Ukraine;
e-mail: yurap@mail.ru

IDENTIFYING BUSINESS COMPONENTS FROM BUSINESS MODEL: A METHOD BASED ON FEATURE MATCHING

Meng Fanchao, Zhan Dechen, Xu Xiaofei

Abstract: Identifying reusable business components from business model is the premise of Component-Based Software Development (CBSD). In CBSD, business component is the basic unit for reuse and it provides a coarse-grained functionality. A business component typically consists of related elements that possess similar features. This paper proposes an approach to business components identification based on features matching. In our method, the concepts of feature and equivalent feature relation are presented, and the rule of judging equivalent feature relation and the algorithm of parting feature set are given. To identify reusable business components, a hierarchical clustering technique is proposed. In the process of clustering, we give the formula of calculating similarity among a set of elements that extends Sorenson Coefficient. Finally, a tool RBCET is built using this method to help reusable business components extraction from domain business model.

Keywords: domain business model; feature; equivalent feature set; reusable business component.

ACM Classification Keywords: D.2 Software – Software Engineering (K.6.3)

Introduction

With the rapid development of hardware of computer, software has become more and more complex. How to rapidly develop maintainable, extensible and adaptable software that meet the changeable requirements has become a crucial problem. Component-Based Software Development (CBSD) plays an important role in tackling software crisis and promoting the software productivity [D'Souza, 1999] [Jain H, 2001]. In CBSD, component is the basic unit for reuse and it provides a coarse-grained functionality. Identification of reusable components is the

premise of CBSD [yang Fuqing, 1999], currently there are some methods have been brought forward for resolving this problem. In general, we classify those methods into two categories: structure analysis and feature matching.

The methods of identifying reusable business components from domain business model can be classified into two categories: structure analysis methods and feature matching methods. Structure analysis methods abstract domain business model as mathematical notations, such as tree structure or a graph structure. Via cluster analysis, a domain business model can be partitioned into sub-structures, and each sub-structure is taken as a candidate component. Currently the main methods of structure analysis have COMO [Lee SD, 1998], O2BC [Ganesan R, 2001], CRWD[Somjit Arch-int, 2003] and graph decomposition [Y. Chiricota, 2003]. A disadvantage of these methods is that the results of partition excessively depend on weights which are set by designer so that it is difficult to apply into practice. Feature matching methods classify similar elements according to the features of them, and these methods also depend on the similarity measures and clustering algorithms being used. Wiggers [T.A. Wiggers, 1997] gives an overview of software clustering techniques and suggests the use of the term 'entity' to describe elements being grouped together and 'feature' to denote the attributes of these elements. The representative feature matching method is the F³ reuse methodology proposed by AIPA [Silvana Castano, 1997,1998]. In this methodology, reusable conceptual components are constructed from schema families stored in the Design Library using descriptors. They calculate the conceptual distance between components in different schemas and cluster them according to similarity levels based on the computation of an affinity measure between components.

In this paper, a method of reusable component identification based on feature matching is proposed. In our approach, the concepts of feature and equivalent feature relation are presented, and give formula of calculating resemble degree among business elements which extends Sorenson Coefficient. As our experiments show, the method proposed can provide more promising results for component identification than the previously used methods.

Domain Business Model and Business Components

A domain business model is composed of a set of business models that belong to same application domain. A business model consists of a business object model and a business process model.

A business object model is composed of a set of business objects and relationships between them. A business object is an object with well defined boundaries and an identity that encapsulates a business state and behavior [13]. A business state is a structure property represented by attributes while a behavior is a behavioral property represented by business operations that operate on the attributes. Business objects represent resources in a business model such as product, planning, order, material etc.

Definition 1: A business object can be defined as $BO = \{n, A, OP, R\}$, where n is name of business object, A is the set of attributes, and OP is the set of business operations that operate on attributes. R is the set of relationships between the business objects and other business objects.

Attribute can be classified into individual attribute and composite attribute. A individual attribute is defined with a name and a data type, $a = \{n, DT\}$, where n represents the name of attribute, DT represents the data type of attribute. A composite attribute is defined groups of individual attributes logically related and grouped, denote as $a = (a_1, a_2, \dots, a_n)$, where $a_i (i=1, 2, \dots, n)$ is a individual or composite attribute.

Business object can provide business activities with business operations to satisfy their executive demands. A business operation is defined as $op = \{n, t, In, Out\}$, where n is the name of business operation. $t \in \{Create, Modify, Delete, Transform, Query\}$ is the type of business operation. In is the set of input parameters and Out is the set of output parameters. A parameter can be represented as $p = (n_p, DT_p)$, where n_p is the name of parameter, and DT_p is the data type of parameter.

A relationship between two business objects BO_1 and BO_2 can be defined as $r_{12}=(n_1, n_2, t)$, where, n_1 is the name of BO_1 , n_2 is the name of BO_2 , $t \in \{Generalization, Association, Dependency\}$ is the type of relationship. A generalization is a taxonomic relationship between a general business object BO_1 and a specific business object BO_2 . An association specifies a link relationship that can occur between BO_1 's instances and BO_2 's instances. A dependency is relationship that signifies that one or more business operations in BO_1 call the business operations in BO_2 for their implementation.

A business process model can be decomposed into a set of business processes and each business process can also decomposed into business activities resulting in a two-level hierarchy the business processes model. A business process is a specific ordering of business activities across time and place, with a beginning, an end, and clearly identified inputs and outputs. These ordered business activities affect the states of business objects by creating, consuming and changing their contents. Business activity that involves business objects operating on a business state with business operation is the basic function unit of a particular business process.

Definition 2: A business activity can be defined as: $BA=\{n, In, Out, OP\}$, where n is name of business activity. In is the set of input business objects, and Out is the set of output business objects. OP is the set of business operations, and each business operation is provided by corresponding business object.

Traditionally, a software component is defined as a self-contained piece of software with well-defined interface or set of interfaces [14]. A larger-grained component called a business component focuses on a business concept as the software implement of an autonomous business concept or business process [G.Q. Huang, 1999]. Business components vary from traditional software artifacts such as code segment, class and procedure etc. Traditional software artifacts are mostly fine-grained and technical-oriented, business components, on the other hand, are more coarse-grained and provide a high-level business-oriented representation, and they can express future components and the relations of those components.

In term of the functions implemented by business components, they can be classified into: entity components and process component. In general, business objects that possess resemble features in a domain business model are capsulated an entity component. Analogously, business activities that carry out resemble tasks in a domain business model are capsulated a process component.

Equivalent Feature Relations

To identify reusable business components from domain business models, we use the elements represent business objects and business activities in domain business model, and use the features represent the characteristics of business objects and business activities. According to the definitions of business object and business activity, the features of business object include name, attributes, business operations and relationships, and the features of business activity include name, input business objects, output business objects and business operations.

To evaluate similarity between elements in different business models in a domain business model, we need to refer to the domain thesaurus containing semantic information. A thesaurus usually is sets of dictionaries, every one of which contains group of terms that are extracted the names of business elements (business objects, attributes, date type, business operation and business activities, etc) from all business models in a domain business model. Each dictionary in domain thesaurus is structured as a directed graph. Nodes of the graph represent the terms and directed edges between nodes represent the partial relations between terms. The distance between two terms reflects the semantic similarity between them. The longer the distance is, and the less the similarity is. In the following, we give the definitions of similarity relation between features.

- **Name Similarity**

Let n_1 and n_2 be two business objects' names, if $SIM(n_1, n_2) \geq \theta_{BO}$, where, $\theta_{BO} (0 \leq \theta_{BO} \leq 1)$ is a similarity threshold of the names of business objects, then n_1 and n_2 are similar, denoted as $n_1 \sqsim n_2$.

Let n_1 and n_2 be two business activities' names, if $SIM(n_1, n_2) \geq \theta_{BA}$, where, $\theta_{BA} (0 \leq \theta_{BA} \leq 1)$ is a similarity threshold of the names of business activities, then n_1 and n_2 are similar, denoted as $n_1 \sqsim n_2$.

- **Attribute Similarity**

Let $a_1 = (n_1, DT_1)$ and $a_2 = (n_2, DT_2)$ be two individual attributes, if a_1 and a_2 satisfy condition: $SIM(n_1, n_2) \geq \theta_A$, where, $\theta_A (0 \leq \theta_A \leq 1)$ is a similarity threshold of attribute names, then a_1 and a_2 are similar, denoted as $a_1 \sqsim a_2$.

Let $f = (a_1, a_2, \dots, a_n)$ and $g = (b_1, b_2, \dots, b_m)$ be two composite attributes, if there exists a permute function T such that $T(a_1, a_2, \dots, a_n) = (a'_1, a'_2, \dots, a'_n)$ and $T(b_1, b_2, \dots, b_m) = (b'_1, b'_2, \dots, b'_m)$. If f and g satisfy condition: $(m=n) \wedge (a'_1 \sqsim b'_1) \wedge (a'_2 \sqsim b'_2) \wedge \dots \wedge (a'_n \sqsim b'_m)$, then f and g are similar, denoted as $f \sqsim g$.

- **Business Operation Similarity**

Let $p_1 = (n_1, DT_1)$ and $p_2 = (n_2, DT_2)$ be two parameters, if p_1 and p_2 satisfy condition: $SIM(n_1, n_2) \geq \theta_P$, then p_1 and p_2 are similar, denoted as $p_1 \sqsim p_2$. Let $P_1 = \{p_{11}, p_{12}, \dots, p_{1m}\}$ and $P_2 = \{p_{21}, p_{22}, \dots, p_{2n}\}$ be two set of parameters,

$$SIM(P_1, P_2) = \frac{2 \cdot |P(P_1 \cap P_2)|}{|P_1| + |P_2|}, \text{ where, } P(bc_1 \cap bc_2) = \{(p, p') \mid p \in P_1, p' \in P_2, p \sim p'\}.$$

Let $bop_1 = (n_1, t_1, In_1, Out_1)$ and $bop_2 = (n_2, t_2, In_2, Out_2)$ be two business operations, if bop_1 and bop_2 satisfy condition: $SIM(n_1, n_2) \geq \theta_{BOP} \wedge (t_1 = t_2) \wedge SIM(In_1, In_2) \geq \alpha \wedge SIM(Out_1, Out_2) \geq \beta$ (θ_{BOP} is a similarity threshold of the names of business operations, α is a similarity threshold of input parameters, and β is a similarity threshold of output parameters, $0 \leq \beta, \gamma, \theta \leq 1$), then bop_1 and bop_2 are similar, denoted as $bop_1 \sim bop_2$.

- **Relationship Similarity**

Let $r = (n_1, n_2, t)$ and $r' = (n'_1, n'_2, t')$ be two relationships between business objects, if r and r' satisfy condition: $SIM(n_1, n_2) \geq \theta_{BO} \wedge SIM(n'_1, n'_2) \geq \theta_{BO} \wedge (t = t')$, then r and r' are similar, denoted as $r \sqsim r'$.

Based on the similarity relations between features, we can define the equivalent relations between features. Let F is a set of finite features, f_1 and f_2 be two features on set F ,

- If $f_1 \sim f_2$, then f_1 and f_2 have equivalence relation, denoted as $f_1 \cong f_2$.
- Equivalent feature relation is transitive, that is to say, if $f_1 \cong f_2, f_2 \cong f_3$, then $f_1 \cong f_3$.

Let F be a feature set, for every feature $f \in F$, equivalence feature set of f is defined as: $[f] \cong = \{f' \mid (f' \in F) \wedge (f' \cong f)\}$, and the partition on set F can be defined as: $F/\cong = \{[f] \cong \mid f \in F\}$.

Let DBM be a domain business model, $BOS = \{BO_1, BO_2, \dots, BO_m\}$ be the set of business objects in DBM , $BAS = \{BA_1, BA_2, \dots, BA_n\}$ be the set of business activities in DBM . In the following, we give some symbols:

- $N(BOS)$: the set of names of all business objects in BOS , $N(BOS)/\cong = \{[n] \cong \mid n \in N(BOS)\}$.
- $A(BOS)$: the set of attributes of all business objects in BOS , $A(BOS)/\cong = \{[a] \cong \mid a \in A(BOS)\}$.
- $OP(BOS)$: the set of business operations of all business objects in BOS , $OP(BOS)/\cong = \{[op] \cong \mid op \in OP(BOS)\}$.
- $R(BOS)$: the set of relationships between business objects in BOS , $R(BOS)/\cong = \{[r] \cong \mid r \in R(BOS)\}$.
- $N(BAS)$: the set of names of all business activities in BAS , $N(BAS)/\cong = \{[n] \cong \mid n \in N(BAS)\}$.
- $IN(BAS)$: the set of input business objects' name of all business activities in BAS , $IN(BAS)/\cong = \{[n] \cong \mid n \in IN(BAS)\}$.
- $Out(BAS)$: the set of output business objects' name of all business activities in BAS , $Out(BAS)/\cong = \{[n] \cong \mid n \in Out(BAS)\}$.
- $OP(BAS)$: the set of business operations of business activities in BAS , $OP(BAS)/\cong = \{[op] \cong \mid op \in OP(BAS)\}$.

To acquire equivalence feature set in domain business model, we give the algorithm of parting feature set which are described as following:

Algorithm 1: partition of equivalence feature set

Input: $F = \{f_1, f_2, \dots, f_n\}$;

Output: $F/\cong = \{[f]_{\cong} \mid f \in F\}$;

```

1   $TF \leftarrow F; F/\cong \leftarrow \phi$ ;
2  for ( each  $f_k \in TF$  )
3  {
4    Add ( $[f_k]_{\cong}, f_k$ );
5    Remove ( $TF, f_k$ );
6    for ( each  $f_j \in TF$  )
7    {
8      if ( $f_j \cong f_k$ )
9      {
10       Add ( $[f_k]_{\cong}, f_j$ );
11       Remove ( $TF, f_j$ );
12     }
13   }
14   Add ( $F/\cong, [f_k]_{\cong}$ );
15 }
```

The functions used in the algorithm are defined as follows:

- Add ($[f_k]_{\cong}, f_k$) add element f_k into set $[f_k]_{\cong}$.
- Remove (TF, f_k) delete element f_k from set $[f_k]_{\cong}$.

According to the algorithm, if we input $N(BOS)$ (resp. $A(BOS), OP(BOS), R(BOS), N(BAS), In(BAS), Out(BAS)$ and $OP(BAS)$), it can output $N(BOS)/\cong$ (resp. $A(BOS)/\cong, OP(BOS)/\cong, R(BOS)/\cong, N(BAS)/\cong, In(BAS)/\cong, Out(BAS)/\cong$ and $OP(BAS)/\cong$).

Similarity among a set of elements

A similarity for a given pair of elements indicates the degree of resemblance between the two elements. The metrics to calculate similarity between two elements have Jaccard coefficient, Sorensen coefficient, Russel and Rao coefficient, simple matching coefficient, Soka and sneath and Yule coefficient, etc. An approach may be well suited to one domain but not to another. For the identification of business components, the Jaccard coefficient and Sorensen coefficient metrics are more appropriate than others [S.Mancoridis, 1999]. In this paper, we extend Sorensen coefficient to calculate similarity among a set of elements. Different to the method followed by Davey and Burd [J.Davey, 2000], we use business objects and business activities as elements, and use equivalence feature sets as the attributes of the elements. In this paper, we use equivalence feature matrix calculate the similarity between a set of elements.

An equivalence feature matrix can be defined as $M=[E, F/\cong]=[m_{ij}]_{m \times n}$, where $E=\{e_1, e_2, \dots, e_m\}$ is the set of elements, F is the set of features that belong to the elements in E , and $F/\cong=\{[f_1]_{\cong}, [f_2]_{\cong}, \dots, [f_n]_{\cong}\}$ is the set of equivalence feature sets on F . If there exists a feature $f \in F(e_i)$ ($1 \leq i \leq n$), such that $f \in [f_j]_{\cong}$ ($1 \leq j \leq n$), then $m_{ij}=1$, else $m_{ij}=0$.

Let $E_k = \{e_{k1}, e_{k2}, \dots, e_{ks}\} \subseteq E$ be a subset of elements in E , $[f]_{\Xi} = (m_{1j}, m_{2j}, \dots, m_{nj})$ be a column of matrix M , if $m_{ij} = 1$ for every $i \in \{k1, k2, \dots, ks\}$, then $[f]_{\Xi}$ is called matching type t_1 on $e_{k1}, e_{k2}, \dots, e_{ks}$; if there exists $i \in \{k1, k2, \dots, ks\}$ such that $m_{ij} = 1$, and there also exists p ($p \neq i, i, p \in \{k1, k2, \dots, ks\}$) such that $m_{pj} = 0$, then $[f]_{\Xi}$ is called as matching type t_2 on $e_{k1}, e_{k2}, \dots, e_{ks}$; if $m_{ij} = 0$ for every i ($1 \leq i \leq n$), then $[f]_{\Xi}$ is called matching type t_3 on $e_{k1}, e_{k2}, \dots, e_{ks}$. We denote as $T_1(e_{k1}, e_{k2}, \dots, e_{ks})$, $T_2(e_{k1}, e_{k2}, \dots, e_{ks})$ and $T_3(e_{k1}, e_{k2}, \dots, e_{ks})$ the set of matching type t_1 , t_2 and t_3 on $e_{k1}, e_{k2}, \dots, e_{ks}$. The similarity among $e_{k1}, e_{k2}, \dots, e_{ks}$ can be defined as:

$SIM(e_{k1}, e_{k2}, \dots, e_{ks}) = |T_1(e_{k1}, e_{k2}, \dots, e_{ks})| / (|T_1(e_{k1}, e_{k2}, \dots, e_{ks})| + \sum_{m_j \in T_2(e_{k1}, e_{k2}, \dots, e_{ks})} r(m_j))$, where, $r(m_j)$ represents the proportion of 0 in column $[f]_{\Xi}$. If $n=2$, then $SIM(e_1, e_2) = 2a/(2a+b)$ ($a = |T_1(e_1, e_2)|, b = |T_2(e_1, e_2)|$) which becomes Sorenson coefficient.

Table 1 Feature Matrix

Elements	Equivalence Feature Sets						
	$[f_1]_{\Xi}$	$[f_2]_{\Xi}$	$[f_3]_{\Xi}$	$[f_4]_{\Xi}$	$[f_5]_{\Xi}$	$[f_6]_{\Xi}$	$[f_7]_{\Xi}$
e_1	1	1	0	0	1	1	0
e_2	0	1	0	0	1	0	0
e_3	0	1	1	1	0	0	1
e_4	0	0	1	1	0	0	1
e_5	1	0	1	0	1	0	1

Here, we give an example to explain the method of calculating the similarity among a set of elements. Table 1 gives a feature matrix that consists of five elements and seven equivalence feature sets.

- $T_1(e_3, e_4, e_5) = \{[f_3]_{\Xi}, [f_4]_{\Xi}\}$, $T_2(e_3, e_4, e_5) = \{[f_1]_{\Xi}, [f_2]_{\Xi}, [f_4]_{\Xi}, [f_5]_{\Xi}\}$, $T_3(e_3, e_4, e_5) = \{[f_6]_{\Xi}\}$, $SIM(e_3, e_4, e_5) = 2/(2 + (2/3 + 2/3 + 1/3 + 2/3)) = 6/13$.
- $T_1(e_1, e_2, e_3, e_4, e_5) = \emptyset$, $SIM(e_1, e_2, e_3, e_4, e_5) = 0$.

Let DBM be a domain business model, BOS be the set of business objects in DBM , BAS be the set of business activities in DBM .

The similarity among a subset of business objects $BOS_i = \{BO_{i1}, BO_{i2}, \dots, BO_{ik}\} \subseteq BOS$ can be defined as: $SIM(BOS_i) = w_N \cdot SIM_N(BOS_i) + w_A \cdot SIM_A(BOS_i) + w_{OP} \cdot SIM_{OP}(BOS_i) + w_R \cdot SIM_R(BOS_i)$, where, $SIM_N(BOS_i)$ is the name similarity among $BO_{i1}, BO_{i2}, \dots, BO_{ik}$, $SIM_A(BOS_i)$ is the attribute similarity among $BO_{i1}, BO_{i2}, \dots, BO_{ik}$, $SIM_{OP}(BOS_i)$ is the business operation similarity among $BO_{i1}, BO_{i2}, \dots, BO_{ik}$, and $SIM_R(BOS_i)$ is the relationship similarity among $BO_{i1}, BO_{i2}, \dots, BO_{ik}$. w_N (resp. w_A , w_{OP} and w_R) is the weight of name (resp. attribute, business operation and relationship), $w_N + w_A + w_{OP} + w_R = 1$, $0 \leq w_N, w_A, w_{OP}, w_R \leq 1$. we can set

The similarity among a subset of business activities $BAS_i = \{BA_{i1}, BA_{i2}, \dots, BA_{js}\} \subseteq BAS$ can be defined as: $SIM(BAS_i) = w_N \cdot SIM_N(BAS_i) + w_{In} \cdot SIM_{In}(BAS_i) + w_{Out} \cdot SIM_{Out}(BAS_i) + w_{OP} \cdot SIM_{OP}(BAS_i)$, where, $SIM_N(BAS_i)$ is the name similarity among $BA_{i1}, BA_{i2}, \dots, BA_{js}$, $SIM_{In}(BAS_i)$ is the input similarity among $BA_{i1}, BA_{i2}, \dots, BA_{js}$, $SIM_{Out}(BAS_i)$ is the output similarity among $BA_{i1}, BA_{i2}, \dots, BA_{js}$, and $SIM_{OP}(BAS_i)$ is the business operation similarity among $BA_{i1}, BA_{i2}, \dots, BA_{js}$. w_N (resp. w_{In} , w_{Out} and w_{OP}) is the weight of name (resp. input, output and business operation), $w_N + w_{In} + w_{Out} + w_{OP} = 1$, $0 \leq w_N, w_{In}, w_{Out}, w_{OP} \leq 1$.

Business Component Identification Algorithm

To identifying reusable business components from domain business model, we use the hierarchical algorithm to group these elements whose similarity is more than a threshold given to a business component. Hierarchical algorithms start from the individual elements, gather them into small clusters which are in turn gathered into larger

clusters. At each step, the two clusters that are closest to each other are merged and the number of clusters is reduced by one.

Algorithm 2:

Input: $E=\{e_1, e_2, \dots, e_m\}$; θ is the similarity threshold to choose the high reusable business components; s is the size threshold for the result of clustering.

Output: $P(E)=\{E_1, E_2, \dots, E_n\}$ that satisfies condition: (1) $E_1 \cup E_2 \cup \dots \cup E_n = E$ ($n \leq s < m$); (2) $E_i \cap E_j = \emptyset$ ($1 \leq i, j \leq n$; $i \neq j$).

```

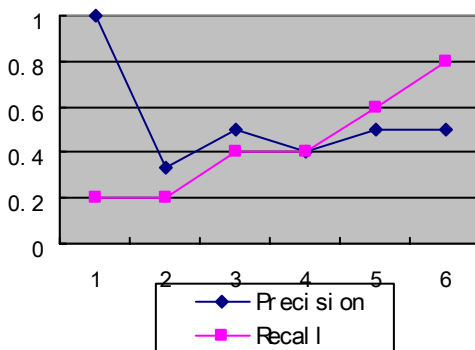
1   $P(E) = \emptyset$ ;
2  for( $i=1$ ;  $i \leq m$ ;  $i++$ ) {
3     $E_i = \{e_i\}$ ;
4    Add( $P(E), E_i$ );
5  }
6  while( $|P(E)| > s$ ) {
7    Chose two clusters  $E_k$  and  $E_j$  from  $P(E)$  such that  $SIM(E_k \cup E_j)$  is the biggest for any two clusters in  $P(E)$ ;
8    if( $SIM(E_k \cup E_j) > \theta$ ) {
9       $E_k = E_k \cup E_j$ ;
10     Delete( $P(E), E_j$ );
11   } else break;
12 }
```

The functions used in the algorithm are defined as follows:

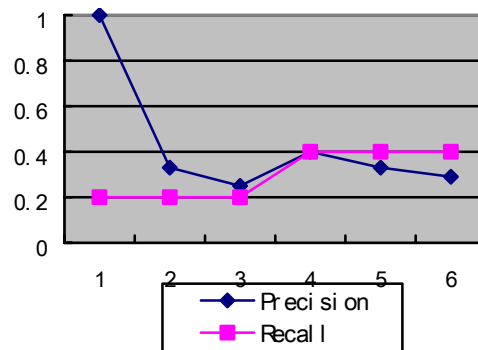
- Add($P(E), E_i$) add element E_i into set $P(E)$.
- Delete($P(E), E_j$) delete element E_j from set $P(E)$.
- According to above algorithm, if we input $BOS=\{BO_1, BO_2, \dots, BO_m\}$, it output the clusters of business objects, each of which can be identified a entity component; if we input $BAS=\{BA_1, BA_2, \dots, BA_n\}$, it output the clusters of business activities, each of which can be identified a process component.

Case and Experiments

In this section, we use inventory management system as example to demonstrate the proposed approach to identify reusable components from domain business model. In this case, we give three business models that represent different business requirements in inventory management domain. To evaluate the result of business components identification, we select some representative business activities from the domain business model. First we perform the component capture manually. Because the complete linkage algorithm can get the best partitions out of single, weighted, unweighted linkage algorithms [J.Davey, 2000], we compare the performance of complete linkage algorithm with that of our approach. As can be seen from figure 1, the precision and recall for the clustering algorithm is higher than form complete algorithm.



(a) Our approach



(b) Complete algorithm

Fig.1: Precision and Recall

Reusable Business Component Extract Tool (RBCET)

To assist the designer to cluster similar business objects and business activities define reusable business components from domain business model, we have design and develop Reusable Business Component Extract Tool (RBCET). Figure 2 shows the process of identifying reusable business components from domain business model using RBCET.

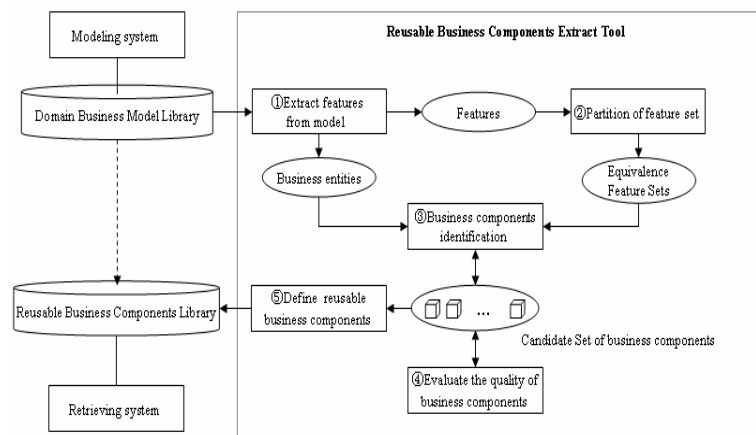


Fig 2: The process of reusable business components Identification

Conclusion

This paper present a method of reusable business components identification based on domain business model and these components are organized in a Library called reusable business component library which can constitute a repository of the core knowledge of an enterprise in a given domain. The proposed method has been experiment on a lot of domain business model. Based on the method, the reusable business component extract tool (RBCET) has been designed and implemented to assist the designer to cluster similar business objects and business activities define reusable business components from domain business model.

Bibliography

- [A.W. Brown, 1998]A.W. Brown, C. Wallnau, The current state of CBSE, IEEE Software 15 (5) (1998) 37–46.
- [D'Souza, 1999] Df. D'Souza, A C Wills. Objects, Components and Frameworks with UML: the Catalysis Approach. Addison-Wesley, 1999.
- [Ganesan R, 2001]Ganesan R, Sengupta S. OSBC: A technique for the design of component-based applications. In: Proceedings of the 39th International Conference and Exhibition on Technology of Object-Oriented Language and Systems. IEEE computer Society Press, 2001, 46~55.
- [G.Q. Huang, 1999]G.Q. Huang, K.L. Mak, Issues in the development and OMG-Business Object Domain Task Force, Business Object Concepts, White paper, January 1999 OMG document: bom/99-01-01.
- [Jain H, 2001] Jain H, Chalimeda N, Ivaturi N, Reddy B. Business component identification - A formal approach. In: Proceedings of the 5th IEEE International Enterprise Distributed Object Computing Conference. Seattle: IEEE Computer Society Press, 2001. 183~187.
- [J.Davey, 2000]J.Davey and E.Burd, "Evaluating the Suitability of Data Clustering for Software Remodularization," *The Seventh Working Conference on Reverse Engineering (WCRE'00)*.2000.
- [Lee SD, 1998] Lee SD, Yang YJ. COMO: A UML-based component development methodology. In: Proceedings of the 6th Asia Pacific software Engineering conference. Takamatsu: IEEE Computer Society Press, 1998. 54~63.
- [P. Herzum, 2000]P. Herzum, O. Sims, Business Component Factory: A Comprehensive Overview of Component-Based development for the Enterprise, Wiley, 2000.

-
- [Silvana Castano, 1997]Silvana Castano, Valeria De Antonellis. Engineering a library of reusable conceptual components. Information and Software Technology, 1997, 35(2): 43 ~ 57.
- [Silvana Castano, 1998]Silvana Castano, Valeria De Antonellis et al. Conceptual schema analysis: Techniques and applications. ACM Trans on Database Systems, 1998, 23 (3) : 286 ~ 333.
- [S.Mancoridis, 1999]S.Mancoridis, B.Mitchell, Y. Chen, and E. Gansner. Bunch: A clustering tool for the recovery and maintenance of software system structures. In *Proceedings of the International Conference on Software Maintenance*. IEEE Computer Society Press, 1999.
- [Somjit Arch-int, 2003]Somjit Arch-int, Dentcho N. Batanov. Development of industrial information systems on the Web using business components. Computer in Industry 2003,50(2):231~250.
- [T.A. Wiggerts, 1997]T.A. Wiggerts, "Using clustering algorithms in legacy systems remodularization," *Fourth Working Conference on Reverse Engineering (WCRE'97)*, October, 1997.
- [Y. Chiricota, 2003]Y. Chiricota, F. Jourdan, and G. Melancon, "Software Components Capture using Graph Clustering," *Proceedings of 11th IEEE International Workshop on Program Comprehension*, 10-11 May 2003, 217-226.
- [Yang Fuqing, 1999] Yang Fuqing ,Mei Hong ,Li Keqin. Software reuse and software component technology . Acta Electronica Sinica, 1999, 27(2): 68~75 (in Chinese).
-

Authors' Information

Meng Fanchao, Zhan Dechen and Xu Xiaofei- School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China; e-mail: Fanchao_meng@yahoo.com.cn

APPLICATION OF INTELLECTUAL TECHNOLOGIES IN PROFIT CONTROL

Inna Khmelnitskaya, Oleg Kosenkov

Abstract: *The modern level of IT-progress achievements predetermines transformation of management methods in the enterprise. Profit, being one of the main signs of the efficiency of the company's activity, accumulates its effectiveness. The authors (of this article) suggest «a matrix of profit control» in which approaches of economic analysis ABC and XYZ are used together with one of the most effective instruments in decision making on the basis of information technologies – Data Mining – in order to solve the problem production conjugacy with the use of association task. Having divided the whole production into profitability groups, and then, according to the degree of risk for the company's financial condition, into matrix points, it is possible to reveal the existing problems and their causes in terms of both types of products and places of their origin. In the present article we suggest a solution of this task by means of intellectual analysis, which will allow the company to achieve considerable competitive advantages.*

Keywords: *ABC and XYZ analysis, matrix of profit control, association, algorithms of associative rules searching, rule of anti-monotony, tree of solutions.*

Introduction

Profit is the driving force of the market and it determines three crucial interrelated aspects of the company's activity: what, how and for whom to produce. The reward for entrepreneurial activity, evinced as the profit, will depend on how correctly the company managers will answer the above questions. The world has numerous tools of management of various economic objects, such as, for example, ABC and XYZ analysis. The ABC-analysis allows to determine the contribution of certain goods, a group of goods or clients into the company sales and profit, and to classify them according to the importance degree and the necessity of control, depending on the

volume and the sum of inflows. XYZ (frequency) analysis is intended for the evaluation of the company sales stability, for the detection of those clients who are not included into the number of clients making high finances or those not very expensive goods, the purchases of which are stably not too big in quantity but still are very frequent. Using the results of the XYZ-analysis the company will be able to differentiate the goods nomenclature depending in the stability of the profit got and the forecast sales. In our opinion, the optimal and a frequently used means of decision making is the ABC-XYZ-analysis, which was used for the creation of the “matrix of profit control”, represented in table 1 [1, 2].

Risk	Profit		
	A ($\geq 75\%$)	B ($\geq 20\%$)	C ($\geq 5\%$)
X (up to 33%)	AX	BX	CX
Y (from 33% to 66%)	AY	BY	CY
Z (over 66%)	AZ	BZ	CZ

Table 1 - Matrix of profit control

The following groups A, B and C are shown as the columns:

Group A – production bringing about 75% of the total profit to the company,

Group B – production bringing about 20% of the total profit to the company,

Group C – production bringing about 5% of the total profit to the company,

Groups of risk X, Y and Z: are represented as rows.

X – group of production with minimal risk, i.e. the volume of sales dynamics is positive, the final economic indexes are good.

Y – group of production with middle risk, i.e. the volume of sales dynamics is not stable or is absent; the final economic indexes are satisfactory.

Z – group of production with maximum risk, i.e. the dynamics is negative, the indexes are marginal.

In this case the risk is calculated from the value of the profit variation index. In accordance with the integrated ABC-XYZ-analysis and the applied bases of production rules knowledge, shown in table 2, the non- structured data are classified into nine groups with different characteristics, which to a big extent simplifies work with them [2].

Table 2 – Original bases of knowledge of production rules of distributing of the production into the cells of the profit control matrix.

Rules	Formulation
1	If the profit is $\geq 75\%$, then if the risk is $\leq 33\%$, then «AX» Otherwise if the risk is $\in (33\%; 66\%]$, then «AY» , otherwise if the risk is $> 66\%$, then «AZ»
2	If the profit is $\geq 20\%$, then if the risk is $\leq 33\%$, then «BX» Otherwise if the risk is $\in (33\%; 66\%]$, then «BY» , otherwise if the risk is $> 66\%$, then «BZ»
3	If the profit is $\geq 5\%$, then if the risk is $\leq 33\%$, then «CX» Otherwise if the risk is $\in (33\%; 66\%]$, then «CY» , otherwise if the risk is $> 66\%$, then «CZ»

For example, the goods bringing the company the maximum profit and most stable in the volume of sales will go to group AX, and the low-profit goods with unstable sales will go to group CZ. There is also a possibility to sort and filter any table column, which gives significant convenience in work with the table to the analyst. That means, pressing “Enter” on any cell will result in showing the contents of the corresponding group. The goods here are

also ordered according to their rating (ABC) and variation index (XYZ), and the matrix in detail, for example in cell AX, shown in figure 1, will look as follows:

	A 1	A 2	A 3
X1	Production of group A1X1 Profit in group A1X1	Production of group A2X1 Profit in group A2X1	Production of group A3X1 Profit in group A3X1
X2	Production of group A1X2 Profit in group A2X1	Production of group A2X2 Profit in group A2X2	Production of group A3X2 Profit in group A3X2
X3	Production of group A1X3 Profit in group A1X3	Production of group A2X2 Profit in group A2X3	Production of group A3X3 Profit in group A3X3

Figure 1 – Matrix of profit control in detail.

From the results of the analyses represented in the matrix, one can determine and regularly review the rules and norms of promotional work and finding out the narrow points of certain goods, work with clients, market segments and other nomenclature. This is an irreplaceable tool to improve the efficiency of the system of goods circulation, increase the profit and, consequently, to improve the financial state of the company.

Nevertheless, creation of the matrix will not either divide the goods into “good” and “bad”, or reveal the goods, which should be immediately withdrawn from the list of sales. An additional analysis is always necessary. For example, there is a problem of conjugation of different types of the production sold, i.e. in category BY one can often find associated goods, just like ski- boots for the skies sold. They bring lower profit and are purchased not so regularly by the clients. In category CZ one can often find ski bindings, which bring low profit and are nor purchased regularly. But these goods must be available, otherwise the client, having not found them may not come again to this shop.

The authors solved this problem by means of the task of associations. The application of this task was also caused by a number of objective factors. Modern databases have huge volumes reaching giga- and terabytes, and still there is a tendency for them to increase constantly. This fact causes a stable growth of interest to the methods of finding out certain knowledge in databases. The volumes of modern databases or information depositories being very impressive created stable demand for the scaled algorithms of the data analyses. One of the popular methods of information finding is algorithms of associative rules searching [3, 4].

The first algorithm of associative rules searching called AIS was worked out in 1993 by the staff of IBM Almaden Research Center [5]. Starting from this pioneer work, the interest to association rules increased; in the middle of the 90-s of the previous century there was the peak of the researches in this field, and since then a few algorithms appear every year. Algorithm revelation of association rules appeared as one of the variants to find out typical purchasing patterns, made in supermarkets. Our task is very correlative (practically the same) with that one, to solve which the principal of revelation of associative dependencies in big databases was worked out. The rules, generated by this algorithm, can be presented graphically and in general can be shown as follows:

- If the customer bought goods A_1 , goods $A_2 \dots$ and goods A_n , he is very likely to buy goods B_1 , goods $B_2 \dots$ and goods $B_m \dots$
- If the customer bought goods A_1 , goods $A_2 \dots$ and goods A_n , he is very likely to buy goods C_1 , goods $C_2 \dots$ and goods $C_k \dots$
- If the customer bought goods D_1 , goods $D_2 \dots$ and goods D_t , he is very likely to buy goods E_1 , goods $E_2 \dots$ and goods E_p .

Here $(A_1, A_2, \dots, A_n), \dots, (D_1, D_2, \dots, D_t)$ – are those goods, that the customer has already bought in the current transaction, and $(B_1, B_2, \dots, B_m), \dots, (C_1, C_2, \dots, C_k), \dots, (E_1, E_2, \dots, E_p)$ – are the goods, that the customer is inclined to buy during the same visit.

Let there is a database consisting of buyers' transactions. Each transaction is the totality of goods bought by the buyer during one visit. This kind of transaction is also called a market basket.

Let $I = \{i_1, i_2, i_3, \dots, i_n\}$ is the multitude (range) of goods called elements. Let D is the multitude of transactions where each transaction T is a set of elements from I : $T \subseteq I$. Each transaction is a binary vector, where $t[k] = 1$, if the element i_k is present in the transaction, otherwise $t[k] = 0$. We say that transaction T includes X , a certain set of elements from I , if $X \subseteq T$. Associative rule is the implication $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ has the support – s , if s % of transactions from D contain $X \cup Y$, $\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y)$. The confidence of the rule shows to what extent it is possible that Y follows from X . The rule $X \Rightarrow Y$ is true with the confidence c , if c % of transaction from D containing X , contains Y too:

$$\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X).$$

In other words, the analysis objective is ascertainment of the following dependences: if in the transaction there is a range of elements X , on these grounds it is possible to conclude that the other range of elements Y must appear in this transaction too. The ascertainment of this kind of dependences allows us to find very simple and intuitively clear rules.

Algorithms of search for associative rules are assigned to find all the rules $X \Rightarrow Y$, for all this support and confidence of these rules must be higher than some thresholds defined beforehand, which are called minsupport and minconfidence respectively.

The search of associative rules is not a trivial task at all, as it may seem from the first glance. One of the problems is algorithmic difficulty at finding frequently met sets of elements, because with the growth of the number of elements in I ($|I|$), there is an exponential growth of the number of potential sets of elements.

The number of columns in the table is equal to the number of elements, present in the set of transactions D . Each note corresponds to a transaction, where in the respective column there is 1, if the element is present in the transaction, and 0 if otherwise. The necessary condition is the fact that the data should be transformed to their normal look, otherwise the algorithm is not applicable.

Nevertheless, to reveal frequently met sets of elements is an operation, demanding a lot of calculation resources and, consequently, time. The primitive approach to solving this task is a prime enumeration of all the possible sets of elements. This will demand $O(2^{|I|})$ operations, where, $|I|$ is a number of elements. Thus appears the necessity to use one of the support characteristics, saying: support of any set of elements cannot exceed the minimal support of any of its subsets.

This characteristic is called anti-monotony and serves to lower the size of the search space. If we did not have this characteristic, finding multi-element sets would be practically impossible because of exponential growth of calculations.

The characteristic of anti-monotony can be formulated in another way: with the growth of size of set of element, the support decreases or remains the same. From all the above-said follows, that any k -elementary set will be frequently met only when all its $(k-1)$ -elementary subsets are frequently met.

All the possible sets of elements from I can be represented as an array, starting with empty set, then on the 1-st level there are 1-element sets, on the 2-nd – 2-element sets and so on. On k -level k -element sets are represented, and they are connected with their $(k-1)$ -element subsets.

Let us consider figure 2, illustrating the set of elements $I = \{A, B, C, D, E\}$. Let us suppose that the set of elements $\{A, B\}$ has the support lower than the given threshold, and, consequently they are not frequently met. Then, according to the anti-monotony characteristic, all its supersets are not frequently met, either, consequently, they are truncated. All this branch, starting from $\{A, B\}$, is marked with grey background. Application of this heuristics allows us to reduce to a great extend the search space.

At the first step of the algorithm 1-element frequently met sets are counted. For this purpose it is necessary to pass through all the set of data and calculate support for them, i.e. how many times they are met in the base.

The following steps will consist of two parts: generation of potentially frequently met sets of elements (they are called candidates) and calculation of the support for the candidates.

After all the frequently met sets of elements have been found, one can start the direct generation of rules.

In order to extract a rule from the frequently met set F it is necessary to find all its nonempty subsets. And for each subset s we can formulate the rule $s \Rightarrow (F - s)$, if the confidence of the rule $\text{conf}(s \Rightarrow (F - s)) = \text{supp}(F)/\text{supp}(s)$ is not less than the threshold minconf .

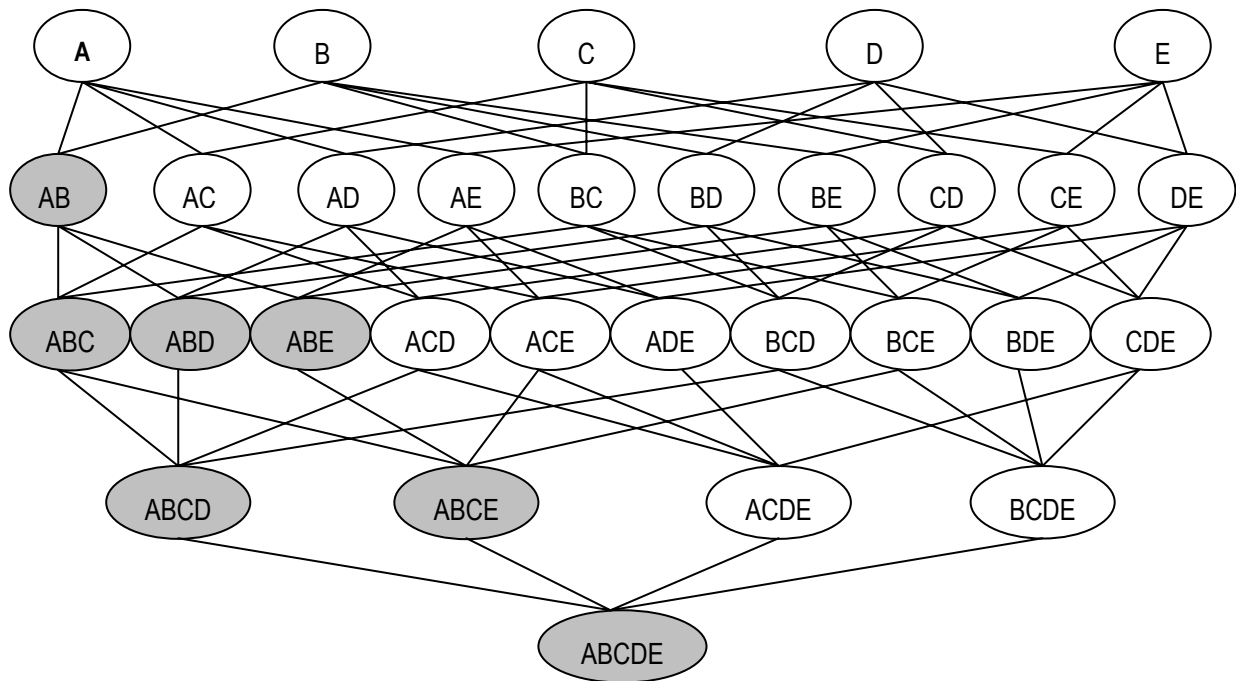


Figure 2 - Graph of application of the rule of anti-monotony at the formation of sets of conjugated production.

Let us comment that the numerator remains constant. The confidence has the minimal value if the denominator has the maximum value, and this happens if in the terms of the rule there is a set, consisting of one element only. All the supersets of the given set have smaller or equal support, and respectively, the higher value of confidence. This property can be used at the rules extraction. If we start to extract the rules, regarding at first only one element in the rule terms, and this rule has the necessary support, then the rules in the terms of which there are supersets of this element, also have the value of confidence higher than the given threshold. For example, if the rule $A \Rightarrow BCDE$ satisfies the minimal threshold of confidence - minconf , then $AB \Rightarrow CDE$ also satisfies it. To extract all the rules recursive procedure is applied. An important comment: any rule, composed of frequently met set, must contain all the elements of the set. For example, if the set consists of the elements $\{A, B, C\}$, then the rule $A \Rightarrow B$ should not be considered.

In this way, applied to our task, it is necessary to establish the level of minimal confidence (minconfidence), the level of which will define the minimal threshold at the selection of the groups of goods (rules), which further on will be considered as one whole at the distribution in matrix cells; in its turn this will allow to take into account the conjugation (mutual dependence) of different types of goods. Still, a problem remains: to define the level of minimal confidence (minconfidence), i.e. if to establish it too high, we may not take into consideration a part of mutually dependant goods, and consequently, not solve the given task, and if we establish a lower level of minimal confidence, we risk to form groups of goods of little dependence on one another, and, consequently, hide

from visual representation the situation with some low-profit goods (having joined them to groups of goods with higher profit). To solve this problem it is possible to use the expert method. As a result of solving a set of tasks (about 20 experiments) it was found out, that the optimal one is the level within 90 up to 100%, depending on the initial quantity of goods. The more it is, the higher should be the level of minimal confidence.

Let us take as an example of one of Belarusian companies to examine the solution of the task considered above. As a result of the analysis of more than 120 kinds of products, the task of association was solved, and owing to this some goods were substituted for the goods-set. That means the distribution of the production into groups A, D and C was made not only for the production which gives a certain share of profit, but also for the acquired with the probability over 90%. The normalized value of mean-square of profit deviation (σ) was used as the risk indicator. At σ value of not more than 0,33 the production was placed to subgroup X, at σ value from 0,33, to 0,66 the production was placed to subgroup Y, at σ value over 0,66 the production was placed to group Z. Thus, the matrix was build, in each cell of which one can find the sum of profit and the group of production placed in this cell (figure 3).

Before the association carrying out			
	A	B	C
X	A, C 7133	E, G 627	N, S, W, U 215
Y	B 868	F, Q, H 484	X, K, L, T, Z 143
Z		P, R, Y 412	D, I, O, M, V 113

After the association carrying out			
	A	B	C
X	C 3060	GN, E 667	S, W, U 175
Y	AFD 4247	QI, H 348	X, K, T, Z 104
Z	BO 901	PRL, Y 451	M, V 42

Figure 3 – Profit control matrixes of one of Belarusian companies before and after the association carrying out

Then the problem of the choice of the analysis order for the production in the cells arose, i.e. production of which cell should be analyzed first of all, which next, etc. For this purpose quadric matrix of pair alternatives comparison was created, and it was filled with estimates from the priorities scale of T. Saati, based on the expert considerations on the following rule. If one alternative of the pair (K1) is more important than the other one (K2) on both criteria of the profit sum and risk, then the cell on the crossing of row K1 and column K2 is filled with the number of expert's evaluation, and the cell of the matrix situated on the crossing of row K2 and column K1- with the reciprocal number. Then, applying one of the possible methods - geometric middling of the line - the vector of priorities is created; in this vector to each row of the matrix corresponds to the meaning of the share of geometric middling value, calculated from the values of each line in the total middling geometric value [8]. The cell with the maximum value will be analyzed first of all and so on according to decrease of values of priority vectors. As a result of practical application within the frameworks of the task solution, one of Belarusian companies got the following vector of priorities (table 3):

Table 3 – Vector of priorities of the analysis of cells of profit control matrix.

Cell	AX	AY	AZ	BX	BY	BZ	CX	CY	CZ
Priority vector	0,391	0,391	0,391	0,391	0,391	0,391	0,391	0,391	0,391
№ of cells analysis	1	2	9	5	4	3	8	6	7

After the order of the analysis is defined, it is necessary to reveal the problems for each group and type of the production, for this purpose one of the methods of Data Mining – the tree of solutions is used [2]. Economic characteristics influencing the profit are shown as nodes, the sign of the branch represents the stimulating direction to profit, thus the volume of sales should be increased (+), and costs should be lowered (–). The tree of factors of profit dynamics formation is represented as a cognitive card in figure 4.

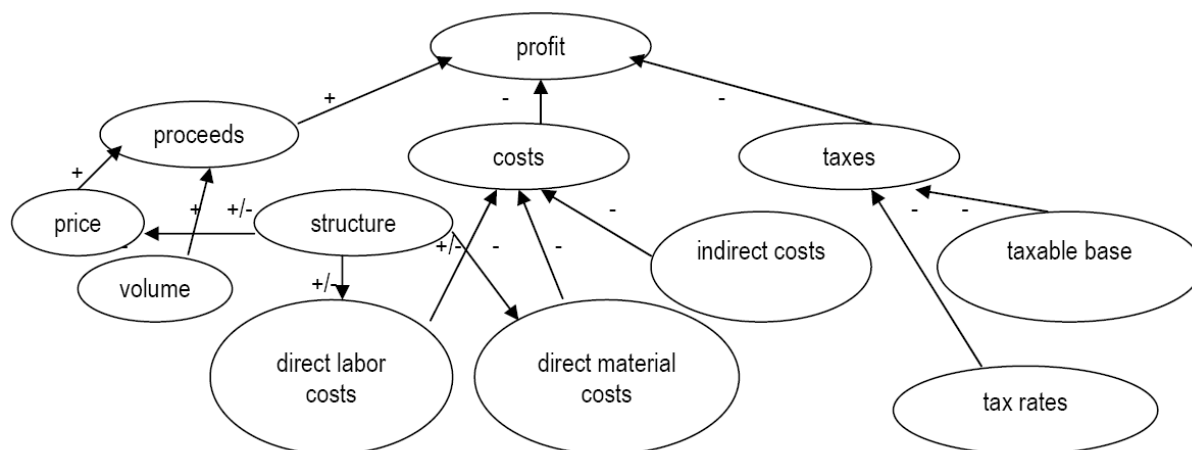


Figure 4 - Cognitive card of interrelation of economic characteristics in their influence on profit

The reason of profit decrease is revealed on the bases of the methodology of marginal analysis application. The critical (maximum permissible value) is calculated for every characteristic (tree node), then the actual value of the characteristic is analyzed in dynamics and compared to the critical value. Thus, for characteristics with negative value, branches of actual value should be lower than critical (maximum permissible) and decrease in dynamics, and for the branches with positive value, the actual value should be higher than the critical (minimal permissible) and increase in dynamics. If the value of the node satisfies the condition of permissibility, then one should move on the following branch. If the condition of permissibility is not observed, the identification of this node takes place- it is identified as the reason of profit decrease and so on and moving on all the branches takes place.

Thus, the matrix of profit control, where such tasks and methods as Data Mining, association, tree of solutions, production rules were realized, will allow the company (depending on the period of time- a day, a month, a year, a few years) to analyze the situation effectively (to reveal the risk degree) both in short-time and long-time periods. This will be realized on the bases of the production classification according to the degree of its meaning for the company financial stability in enlarged groups and in the section of assortment. As a result of the hierarchical structure one can make fragmentation, and the method of the tree of solutions allows to reveal practically all the "narrow" points and, consequently, to take well-grounded management solutions which will allow to greatly improve the actual system of economic analysis on the base of finding out fundamentally new facts and become the basis for considerable competitive advantages.

Bibliography

1. Хмельницкая И.В., Косенков О.А. Новые подходы к анализу и управлению прибылью //Новые информационные технологии в образовании: Доклады и выступления участников седьмой Международной научно-практической конференции «Использование программных продуктов фирмы «1С» в инновационной деятельности учебных заведений» - М.: Федеральное ГОУВПО «Финансовая академия при правительстве Российской Федерации», Фирма «1С», 2007, С.307-310.
2. Хмельницкая И.В., Косенков О.А. Применение Data Mining в экономическом анализе финансово-хозяйственной деятельности предприятия //Инженерный вестник, 2006, №1 (21)/1.- С.192-195.
3. Дюк В., Самойленко А. Data Mining: учебный курс (+CD). –СПб: Питер, 2001. – 368с.:ил.
4. Методы и модели анализа данных OLAP в Data Mining . – СПб., 2004. - 336с.

5. R. Agrawal, T. Imielinski, A. Swami. 1993. Mining Associations between Sets of Items in Massive Databases. In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
 6. S. Brin et al., 'Dynamic Itemset Counting and Implication Rules for Market Basket Data', In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1997.
 7. J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000.
 8. Саати Т. Принятие решений. Метод анализа иерархий. - М.: Радио и связь, 1989. – 316с.
-

Authors' Information

Inna Vladimirovna Khmel'nitskaya – senior lecturer of the Belarusian-Russian University, the town of Mogilev, the Republic of Belarus; postgraduate student of the Belarusian State University, the city of Minsk, the Republic of Belarus; e-mail: inna21@gmail.com

Oleg Alekseevich Kosenkov – senior lecturer of the Belarusian-Russian University, the town of Mogilev, the Republic of Belarus; undergraduate of the Belarusian State University, the city of Minsk, the Republic of Belarus; e-mail: kos.oleg@gmail.com

DEVELOPMENT OF DATABASE FOR DISTRIBUTED INFORMATION MEASUREMENT AND CONTROL SYSTEM

Sergey Kiprushkin, Sergey Kurskov, Vadim Semin

Abstract: The purpose of this work is the development of database of the distributed information measurement and control system that implements methods of optical spectroscopy for plasma physics research and atomic collisions and provides remote access to information and hardware resources within the Intranet/Internet networks. The database is based on database management system Oracle9i. Client software was realized in Java language. The software was developed using Model View Controller architecture, which separates application data from graphical presentation components and input processing logic. The following graphical presentations were implemented: measurement of radiation spectra of beam and plasma objects, excitation function for non-elastic collisions of heavy particles and analysis of data acquired in preceding experiments. The graphical clients have the following functionality of the interaction with the database: browsing information on experiments of a certain type, searching for data with various criteria, and inserting the information about preceding experiments.

Keywords: Database of distributed information measurement and control system, database management system Oracle9i, distance learning.

ACM Classification Keywords: H.2.8 Database management: Database Applications

Introduction

Automatization of scientific research and use of software-driven modular electronics significantly simplify experimental work making it less time-consuming and more accurate. However, a database of an experiment appears to be a non-negotiable condition for data logging, ordered storage and user-friendly maintenance.

The aim of this work was to develop a database for distributed information measurement and control system that would implement methods of optical spectroscopy in atomic collisions and plasma physics as well as provide a remote access to its resources across the Intranet/Internet.

Distributed Information Measurement and Control System

A distinct feature of the distributed information measurement and control system is that it allows combining different device interfaces along with their control computers into uniform network functioning on the basis of TCP/IP.

The distributed system is built as a centralized system [Gavrilov et al., 2003] – [Kiprushkin et al., 2006]. The structure scheme is presented in Figure 1.

The system is comprised of the following parts: the communication (central) server; the equipment servers (CAMAC server, GPIB server, the server of Intel MCS-196 microcontrollers, Ethernet devices server and the server of access at GDS-840C digital oscilloscope et al.); the client programs fulfilling the collection, accumulation and processing of information and experiment control; the universal protocol connecting the communication server with the equipment servers; and the extended protocol connecting the communication server with the client programs.

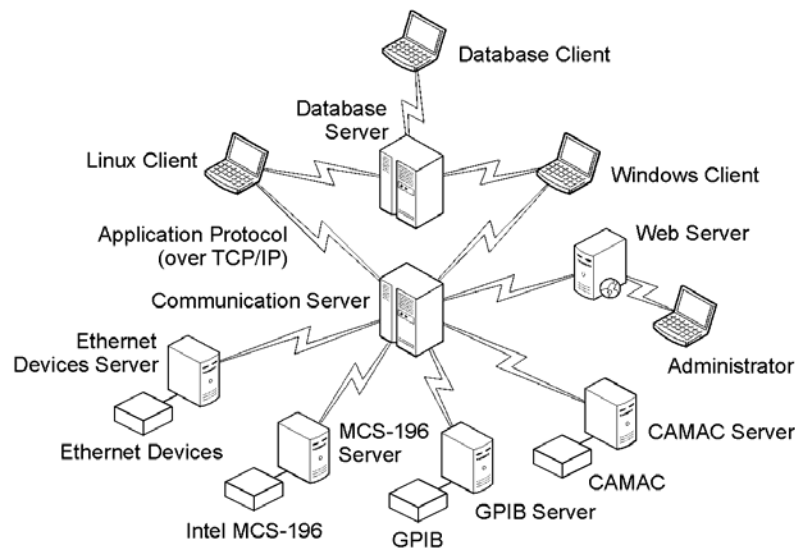


Figure 1. The scheme of the distributed information measurement and control system

The distributed information measurement and control system is based on the modular approach implemented both in the structure and in the software. Clients and equipment servers are built into the system according to the unified rules and interact on a unified protocol by the principles of open systems. Note that an open system is a system that implements open specifications or standards for interfaces, services and formats in order to provide software portability with minimal changes in a wide range of systems (mobility) as well as interaction with other applications on local or remote systems (interoperability) and users (user mobility) [James et al., 1994]. In particular, distributed systems are based on OSE/RM model that describes systems by client/server architecture.

The distributed system was written in Java language – an object-oriented programming language, its platform-independent programs run similarly on diverse hardware with Java virtual machine.

Along with this, web-technologies were widely used, too.

Administration of the distributed system is based on server-side Java servlet. Using the servlet started up on the Web-server, a system administrator logs in to the communication server as a privileged user. S/he monitors resources of the distributed system, may deallocate resources and disconnect clients. The system administrator grants clients' access rights for the equipment servers, ensures unique client and server identifiers, and maintains public-key database of all parts of interconnection.

Note that a communication server of the distributed information measurement and control system is only a moderator between the equipment servers and the client programs that collect, accumulate and process data.

That is, its main function is to maintain multiuser mode and correct allocation of resources among clients, to monitor and to protect the system. Therefore, there is a direct client program-server interaction bypassing the central server. More than that, the latter has no information on the type of the current experiment.

Database of Distributed System

Currently, the most popular database management systems (DBMS) are Borland Interbase, Oracle, MS SQL Server, MS Access and MySQL. Chosen DBMS Oracle9i, namely Intel Pentium 4 compatible Oracle 9.2, perfectly meets such crucial criteria as protection of data integrity, administering capabilities, crossplatformness, guaranteed data storage and recovery, capability to work with large amounts of data, and Java language support [<http://www.oracle.com>; <http://asktom.oracle.com>].

Practically, Oracle DBMS works as a broker between the database and its users. Oracle server is a high-speed multithreaded multiuser SQL server. It can serve high-load critical production systems as well as mass software. The server authorizes clients' access to the database, processes queries and sends information back to clients, or informs them if an accidental error occurs.

When a client sends the data, DBMS verifies if this client is authorized to write the data. If authorization is valid, the server registers it in the database and sends a confirmation back to the client. If access is denied or an error occurs in data write (or transfer), the client receives a notification. It is important that only the core has an access to all data in the database, client applications never write any data directly in the database.

The software for database clients was developed using Model View Controller Architecture (MVS) that separates application data of the model from graphical presentation components and input processing logic (controller).

The following graphical presentations were implemented: measurement of radiation spectra of beam and plasma objects, excitation function for non-elastic collisions of heavy particles and analysis of data obtained in proceeding experiments. Developed client programs manage the course of the experiment by interacting with the equipment servers through the communication server and transferring acquired data to the database. The graphical clients have the following functionality of the interaction with the database: browsing information on experiments of a certain type, searching for data with various criteria, and inserting the information on proceeding experiments.

The developed database stores outputs of the following experiments: measuring optical spectra of plasma objects, excitation spectra of atomic-atomic and ionic-atomic collisions, cross-sections of spectral line excitation with fixed collision energy along with dependence of excitation cross-sections on energy of colliding particles. The data is stored in corresponding tables with Java access. Java Database Connectivity (JDBC) API supports interaction with DBMS Oracle9i. Tables are created using Java utility.

Tables are bound by means of certain relations. This provides the possibility to combine them in a single query. The most frequently used tables are as follows: a table of users and their rights, a table of initial experimental parameters, tables for each type of an experiment where experimental data is logged.

In the distributed system, each part of net intercommunication can login to the database with a unique username and a password. The system administrator registers DBMS users and grants or revokes access at four privilege levels: global, database, table, and column.

The software was implemented using J2SE Development Kit v5.0 and development environment NetBeans 4.0 [<http://java.sun.com>]. A well-designed interface ensures effortless implementation of graphical applications with NetBeans 4.0 tools. A significant advantage of NetBeans is that this environment is produced with the components that allow working with Oracle DBMS using its own API, which provides high-speed data traffic.

Implemented software was assembled into a software complex for Windows operating system with the help of Excelsior JET 3.7 packet. This packet is user-friendly – it automatically installs JRE 1.5, JDBC-Think driver, creates shortcuts on the desktop and in the Programs menu.

Conclusion

The implemented database contributed to the organization of data storage, provided easy and quick access to experimental results, which simplified data processing and analysis of obtained results. The data is available online to researchers from all over the world and can be used as a tool in distance education.

It is necessary to point out that the developed database of distributed information measurement and control system is used for the beam and plasma object analysis with the help of optical spectroscopy methods. In particular, the researches on excitation processes of atomic collisions with inert gas atoms' participation are carried out with its help as well as the laboratory works with senior students of Department of Physics and Engineering of Petrozavodsk State University.

Acknowledgments

We would like to express our gratitude to the laboratories' Head I. P. Shibaev for support of this work as well as post graduate student N. A. Korolev and students M. A. Gvozd, V. G. Mullamekhametov and D. V. Korolev.

Bibliography

- [Gavrilov et al., 2003] S.E.Gavrilov, S.A.Kiprushkin, S.Yu.Kurskov. Distributed information system with remote access to physical equipment. In: Proceedings of the International Conference on Computer, Communication and Control Technologies: CCCT '03 and The 9th International Conference on Information Systems Analysis and Synthesis: ISAS '03. Orlando, Florida, USA, 2003.
- [Kiprushkin et al., 2004] S.A.Kiprushkin, N.A.Korolev, S.Yu.Kurskov. Data security in the distributed information measurement system. In: Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics: SCI 2004. Orlando, Florida, USA, Vol. 1, 2004, pp. 13–16.
- [Kiprushkin et al., 2005] S.A.Kiprushkin, S.Yu.Kurskov, N.G.Nosovich. Resources Control in Distributed Information Measurement System. In: Proceedings of the 3rd International Conference on Computing, Communication and Control Technologies: CCCT '05. Austin, Texas, USA, 2005.
- [Kiprushkin et al., 2005] S.A.Kiprushkin, N.A.Korolev, S.Yu.Kurskov. Sharing of Instrument Resources on the Basis of Distributed Information Measurement System. In: Proceedings of the Second IASTED International Multi-Conference on Automation, Control, and Information Technology - Automation, Control, and Applications: ACIT-ACA 2005. Novosibirsk, Russia, ACTA Press, 2005, pp. 170–175.
- [Kiprushkin et al., 2005] S.Kiprushkin, N.Korolev, S.Kurskov, N.Nosovich. Distributed information measurement system for support of research and education in optical spectroscopy. In: Proceedings of the Third International Conference "Information Research, Applications and Education": i.TECH 2005, Varna, Bulgaria. Ed. Kr.K.Markov. FOI-COMMERCE, Sofia, 2005, pp. 171–179.
- [Kiprushkin et al., 2006] S.Kiprushkin, S.Kurskov, E.Sukharev. Connection of network sensors to distributed information measurement and control system for education and research. In: Proceedings of the XII-th International Conference "Knowledge-Dialogue-Solution": KDS 2006, Varna, Bulgaria. Ed. V.P.Gladun, Kr.K.Markov, A.F.Voloshin et al. FOI-COMMERCE, Sofia, 2006, pp. 347–351.
- [James et al., 1994] I. James et al. Open Systems Handbook. IEEE Standards Press, 1994.

Authors' Information

Sergey Kiprushkin – Petrozavodsk State University, Lenin St., 33, Petrozavodsk -185910, Russia;
e-mail: skipr@dfc3300.karelia.ru

Sergey Kurskov – Petrozavodsk State University, Lenin St., 33, Petrozavodsk -185910, Russia;
e-mail: kurskov@psu.karelia.ru

Vadim Semin – Petrozavodsk State University, Lenin St., 33, Petrozavodsk -185910, Russia;
e-mail: semin@psu.karelia.ru

EXPERT SYSTEM FOR DECISION-MAKING PROBLEM IN ECONOMICS

Aygun Alasgarova, Leyla Muradkhanli

Abstract: *An expert system (ES) is a class of computer programs developed by researchers in artificial intelligence. In essence, they are programs made up of a set of rules that analyze information about a specific class of problems, as well as provide analysis of the problems, and, depending upon their design, recommend a course of user action in order to implement corrections. ES are computerized tools designed to enhance the quality and availability of knowledge required by decision makers in a wide range of industries. Decision-making is important for the financial institutions involved due to the high level of risk associated with wrong decisions. The process of making decision is complex and unstructured. The existing models for decision-making do not capture the learned knowledge well enough. In this study, we analyze the beneficial aspects of using ES for decision-making process.*

Keywords: *expert system, decision-making processing*

1. Introduction

The purpose of this paper is to develop an expert system as an interactive computer program that helps a user solve a problem and make a right decision. Decision making process is inherently complex due to the various forms of risks involved. The existing models for decision-making do not capture the learned knowledge well enough [1-3,5]. In this paper, we analyze the beneficial aspects of using ES for decision-making process and organized as follows: Section 2 provides types of problems in economics solved by ES. Section 3 includes the comparison analysis. Inference rule and confidence are given in Section 4. Section 5 concludes the paper.

2. Types of Problems in economics Solved by ES

Typically, the problems to be solved are of the sort that would normally be tackled by a human "expert" – an economical or other professional, in most cases. Real experts in the problem domain are asked to provide "rules of thumb" on how they evaluate the problems, either explicitly with the aid of experienced system developers, or sometimes implicitly, by getting such experts to evaluate test cases and using computer programs to examine the test data and derive rules from that. Generally expert systems are used for problems for which there is no single "correct" solution which can be encoded in a conventional algorithm — one would not write an expert system to find shortest paths through graphs, or sort data, as there are simply easier ways to do these tasks. Simple systems use simple true/false logic to evaluate data, but more sophisticated systems are capable of performing at least some evaluation taking into account real-world uncertainties, using such methods as fuzzy logic. Such sophistication is difficult to develop and still highly imperfect.

3. ES in Comparison with traditional Problem-solving Systems

The principal distinction between expert systems and traditional problem solving programs is the way in which the problem related expertise is coded. In traditional applications, problem expertise is encoded in both program and data structures. In the expert system approach all of the problem related expertise is encoded in data structures only; none is in programs. Several benefits immediately follow from this organization. An example may help contrast the traditional problem solving program with the expert system approach. The example is the problem of tax advice. In the traditional approach data structures describe the taxpayer and tax tables, and a program in which there are statements representing an expert tax consultant's knowledge, such as statements, which relate information about the taxpayer to tax table choices. It is this representation of the tax expert's knowledge that is difficult for the tax expert to understand or modify. In the expert system approach, the information about taxpayers and tax computations is again found in data structures, but now the knowledge describing the relationships

between them is encoded in data structures as well. The programs of an expert system are independent of the problem domain and serve to process the data structures without regard to the nature of the problem area they describe.

An expert system has four main architectural components that are the knowledge base, the inference engine, the knowledge acquisition module, and the user interface for input/output (Figure 1). The knowledge base contains the domain specific knowledge that is used for problem solving in the domain. Knowledge can be represented and stored in the knowledge base with a format suitable for computer manipulation. One of the most commonly used ways to represent knowledge is as rules (in the form of If-then) [2].

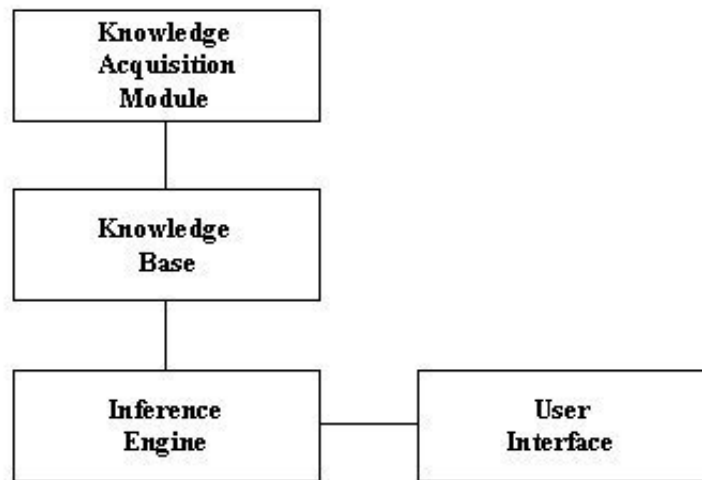


Figure 1. The Architecture of an Expert System

The inference engine is based on an inference rule and a search strategy and contains algorithms. Algorithms are used to manipulate the knowledge stored in the knowledge base in order to solve problems. An inference rule is a way to deduce new knowledge from the existing knowledge base of rules and facts. The knowledge acquisition module enables experts to store their knowledge in the knowledge base or expert system to deduce new knowledge from existing knowledge through a machine learning process. The interface for input/output is used for the expert system to interact with the user, the environment and other systems such as databases, and spreadsheets.

There are generally three individuals having an interaction with expert systems in organization. Primary among these is the end-user; the individual who uses the system for its problem solving assistance. In the building and maintenance of the system there are two other roles: the problem domain expert who builds the knowledge base, and a knowledge engineer who assists the experts in determining the representation of their knowledge and who defines the inference technique required to obtain useful problem solving activity. The end-user usually sees an expert system through an interactive dialog. Knowledge engineers are concerned with the representation chosen for the expert's knowledge declarations and with the inference engine used to process that knowledge.

4. The Inference rule and Confidence

An understanding of the "inference rule" concept is important to understand expert systems. An inference rule is a statement that has two parts, an if-clause and a then-clause.

An expert system's rulebase is made up of many such inference rules. They are entered as separate rules and it is the inference engine that uses them together to draw conclusions. Because each rule is a unit, rules may be deleted or added without affecting other rules (though it should affect which conclusions are reached). One advantage of inference rules over traditional programming is that inference rules use reasoning which more closely resemble human reasoning. Thus, when a conclusion is drawn, it is possible to understand how this conclusion was reached. Furthermore, because the expert system uses knowledge in a form similar to the expert, it may be easier to retrieve this information from the expert.

Actually there are two methods of reasoning when using inference rules: Forward chaining and backward chaining. Forward chaining starts with the available data and uses inference rules to extract more data until an optimal goal is reached. An inference engine using forward chaining searches the inference rules until it

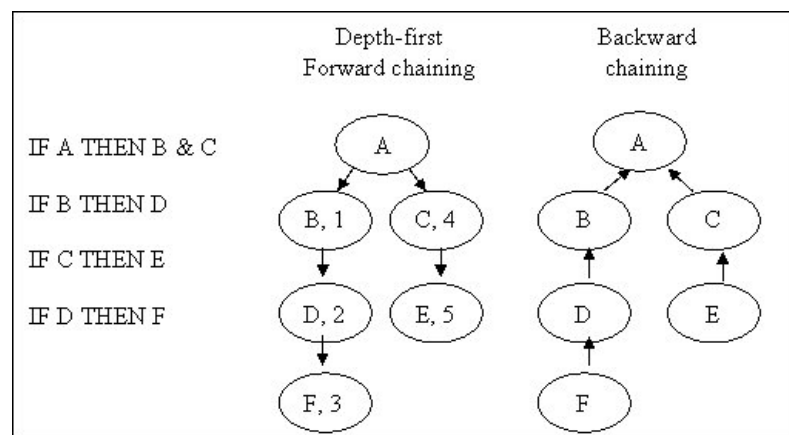


Figure 2. Complex rules

finds one where the If clause is known to be true. When found it can conclude, or infer, the Then clause, resulting in the addition of new information to its dataset. Backward chaining starts with a list of goals (or a hypothesis) and works backwards to see if there are data available that will support any of these goals. An inference engine using backward chaining would search the inference rules until it finds one, which has a Then clause that matches a desired goal. If the If clause of that inference rule is not known to be true, then it is added to the list of goals (in order for your goal to be confirmed you must also provide data that confirms this new rule). The complex rules are demonstrated below in the Figure 2 [6].

Another advantage of expert systems over traditional methods of programming is that they allow the use of confidences. When a human reasons he does not always conclude things with 100% confidence. The user might say, "If the rate is blue, then it is probably increasing". This type of reasoning can be imitated by using numeric values called Confidences. For example, if it is known that rate is blue, it might be concluded with 0.85 Confidence that it is increasing. These numbers are similar in nature to probabilities, but they are not the same. They are meant to imitate the Confidences humans use in reasoning rather than to follow the mathematical definitions used in calculating probabilities [4].

5. Conclusion

The following general points about ES:

- A. The sequence of steps taken to reach a conclusion is dynamically synthesized with each new case. It is not explicitly programmed when the system is built.
- B. ES can process multiple values for any problem parameter. This permits more than one line of reasoning to be pursued and the results of incomplete (not fully determined) reasoning to be presented.
- C. Problem solving is accomplished by applying specific knowledge rather than specific technique. This is a key idea in expert systems technology. It reflects the belief that human experts do not process their knowledge differently from others, but they do possess different knowledge. With this philosophy, when one finds that their expert system does not produce the desired results, work begins to expand the knowledge base, not to re-program the procedures.

So, the ES, including knowledge base, production rule and inference engine, provide a rich and meaningful addition to the traditional methods. This approach is also most likely to be used in a real-world implementation of a decision support system. In this paper we have analyzed the beneficial aspects of using expert system for decision-making process. The kernel of the discussed system is Access DBMS and programming language Visual Basic.

Bibliography

- [1] G.J. Vachtsevanos and S.S. Kim, J.R. Echauz, V.K. Ramani "Neuro-fuzzy approaches to decision making: A comparative study with an application to check authorization". Journal of Intelligent and Fuzzy systems 6 (1998) 259-278.
 - [2] Jinmu Choi "A Rule-Based Expert System Using an Interactive Question-and-Answer Sequence", Dept. of Geography University of Georgia.
 - [3] K. Althoff, S. Wess, and M. Manago, "Reasoning with Cases-Theory and Practice", Proceedings of the European Conference on Artificial Intelligence, pp. 1-26. 1994
 - [4] http://en.wikipedia.org/wiki/Expert_system
 - [5] W. Siler, Birmingham, AL, USA. Building Fuzzy Expert systems. <http://members.aol.com/wsiler/>
 - [6] L. Watson, Applying Case-Based Reasoning, Morgan Kaufmann Publishers, Inc.1997
-

Authors' Information

Aygun N. Alasgarova - Khazar University, PhD student in Computer Engineering and Management; 11 Mehseti St., AZ1096 Baku, Azerbaijan, e-mail: a.alasgarova@mail.ru

Leyla G. Muradkhanli – Khazar University, Assoc. Professor. Sc. Techn.; 11 Mehseti St., AZ1096 Baku, Azerbaijan, e-mail: leyla@khazar.org

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ПРИНЯТИЯ РЕШЕНИЙ И ОБЪЕКТНОГО МОНИТОРИНГА МЕРОПРИЯТИЙ, ПРОВОДИМЫХ НА ТЕРРИТОРИИ РЕГИОНА

**Ирина Петрова, Марина Зайцева, Оксана Щербинина,
Евгений Эрман, Николай Тюрин**

Аннотация: Рассмотрены информационные процессы, связанные с оперативным принятием управленческих решений в задаче контроля за различными мероприятиями, проводимыми на территории региона. Приведены различные категории мероприятий и их основные показатели (географические, финансовые и т.п.). Описаны возможности автоматизированной системы принятия решений и объектного мониторинга мероприятий. Показаны функции компонент, из которых состоит данная автоматизированная система.

Keywords: мероприятие, управленческое решение, объектный мониторинг, категория мероприятия, интернет-компонента, мобильная компонента.

ACM Classification Keywords: J.1Administrative data processing: government

В октябре 2008 года один из областных центров России город Астрахань отмечает свой 450-й день рождения. В рамках подготовки к юбилею определен перечень объектов, подлежащих строительству, капитальному ремонту или реконструкции, запланированы и реализуются организационные мероприятия, связанные с подготовкой и проведением празднования юбилея г. Астрахани.

Качество управленческих решений, оперативность их принятия во многом зависят от актуальности, достоверности и полноты информации об объекте управления и контролируемых объектах. Этот аспект обуславливает необходимость использования в системе управления современных технологий

автоматизации процессов информационной поддержки принятия решений. Таким образом, объектом автоматизации являются процессы сбора, обработки, хранения, анализа, представления и визуализации информации.

Для выявления существенных сторон объекта автоматизации, которые подлежат обязательному учету при формировании требований к функциональности, составу и структуре системы необходимо определение места объекта автоматизации в системе управления.

Управленческим объектом является совокупность органов государственной власти и местного самоуправления (правительство, отраслевые комитеты, ведомственные органы государственной власти совместного подчинения), руководство предприятий и учреждений и служб, дежурные и диспетчерские службы. Под объектом управления понимается совокупность контролируемых объектов.

Исходной информацией, необходимой для принятия управленческих решений, являются:

- достоверные данные о текущем состоянии (местоположении) объектов управления
- критические значения параметров, характеризующих состояние объектов управления
- архивные данные об объектах управления
- данные других информационных систем.

В условиях отсутствия автоматизации процесс сбора информации об объектах является трудоемким и продолжительным. Как правило, информация передается по телефонным (радиотелефонным) каналам, каналам передачи телеметрической информации, либо с использованием бумажных носителей. При этом на достоверность информации существенно влияет субъективный фактор, а ее актуальность по окончании сбора может теряться. Процесс контроля над объектами заключается в сравнении результатов наблюдения (текущих или архивных) с некоторыми критериями, на основании которого вырабатывается заключение о нормальном или критическом состоянии объекта.

Одним из факторов, влияющих на оперативность и качество принимаемых решений, является форма представления информации. Очевидно, что для выработки решений на основе массивов разнотипных количественных и качественных значений показателей необходимо гораздо больше времени, чем в том случае, когда результаты представлены в виде диаграмм, графиков, отражающих существенные стороны процесса. Однако на подготовку данных к представлению также необходимы временные ресурсы.

Необходимость интеграции и тесного регламентированного взаимодействия областных информационных ресурсов в рамках системы поддержки принятия решений представляет собой сложную организационно-техническую проблему. Ее эффективное решение возможно при соблюдении жестких требований в части унификации общесистемных информационных источников; единой картографической основы, справочников и классификаторов улиц и адресных идентификаторов, выверенных данных физических и юридических лиц, объектов недвижимости и прочих городских стационарных и подвижных объектов.

Таким образом, анализ информационных процессов показывает, что необходима их комплексная автоматизация. Только в этом случае лицу, принимающему решение, будет представлена оперативная, актуальная, достоверная и полная информация в удобном для него виде, обеспечена возможность оперативного управления действиями подчиненных сил и средств в масштабе времени, соответствующем сложившейся обстановке.

Силами сотрудников Астраханского государственного университета и Агентства по печати и информационным коммуникациям Астраханской области была создана автоматизированная система принятия решений и объектного мониторинга различных мероприятий. Предлагаемая информационная система предназначена для оперативного принятия управленческих решений и повышения производительности труда сотрудников исполнительных органов государственной власти, связанной со сбором, размещением, оперативным обновлением информации о ходе реализации мероприятий, проводимых на территории региона.

Топологически система поддержки принятия решений и объектного мониторинга представляет собой совокупность локальных узлов (ЛУ), являющихся пунктами концентрации информации, объединенных

одним специального вида ЛУ, реализующим ретрансляцию информационных потоков. Локальные узлы могут представлять собой информационные системы, группы информационных систем, отдельные компоненты систем (БД, клиент-приложение, сервер-приложение) или группы компонент.

Пользователи системы в соответствии с правами доступа получают информацию от ближайшего ЛУ, либо через центр управления доступом - от любого другого локального узла, входящего в состав системы. Средствами локальных узлов обеспечивается единый пользовательский интерфейс в системе. Центр управления доступом обеспечивает единую маршрутизацию, авторизацию и аутентификацию на основе хранящейся в нем информации о пользователях и источниках.

Обобщенная структура системы поддержки принятия решений и объектного мониторинга представлена на рис.1.

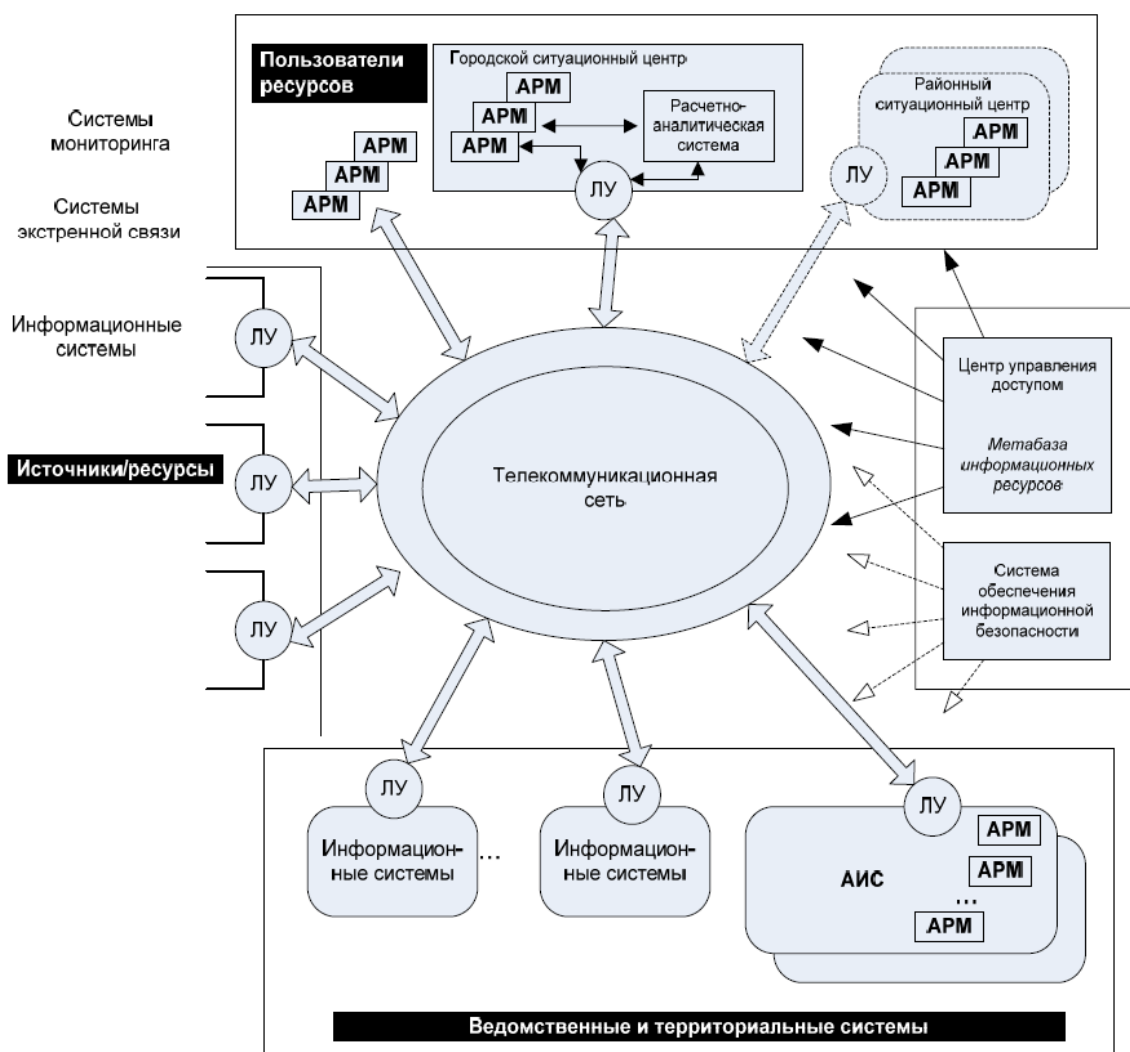


Рис. 1. Обобщенная структура системы поддержки принятия решений и объектного мониторинга выполнения мероприятий

В основу построения информационного обеспечения положены принципы использования первичной документации при создании и эксплуатации информационной базы; одноразовый ввод информации и многократное ее использование; однозначность классификации и кодирования.

Автоматизированная система состоит из следующих компонент:

- подсистемы интерактивной обработки информации по учету и контролю за мероприятиями, проводимыми на территории Астраханской области;

- интернет-компоненты, предназначенной для осуществления контроля за проведением мероприятий всеми заинтересованными лицами;
- мобильной компоненты, предназначенной для работы на карманных переносных компьютерах.

Исполнительные органы государственной власти Астраханской области в соответствии с решениями оргкомитетов, постановлениями и распоряжениями Губернатора и Правительства Астраханской области по подготовке и проведению празднования 450-летия основания г.Астрахани отвечают за выполнение определенных мероприятий. Каждое мероприятие может финансироваться из различных источников. При этом учитывается планируемое финансирование из каждого источника с разбивкой по годам, финансирование на текущий момент (т.е. по факту) и дата завершения подготовки мероприятия. За подготовку и проведение каждого мероприятия отвечает руководитель исполнительного органа государственной власти Астраханской области. Кроме мероприятий существуют также проекты, имеющие межведомственный характер. Такие проекты могут содержать мероприятия, относящиеся к компетенции различных исполнительных органов государственной власти Астраханской области.

В соответствии с планом мероприятий, связанных с подготовкой и проведением празднования 450-летия г.Астрахани, все мероприятия делятся на следующие категории:

- Организационные и культурно-массовые.
- Строительство и реконструкция объектов.
- Реставрация памятников истории и культуры.



Рис.2 Главная форма автоматизированной системы

Каждое из мероприятий, относящихся к категориям «Строительство и реконструкция объектов» и «Реставрация памятников истории и культуры», включает в себя ряд объектов, расположенных в различных частях г.Астрахани, для которых необходимо учитывать ход строительства (реконструкции, реставрации). Объекты, как и мероприятия, могут финансироваться из различных источников, при этом учитывается планируемое финансирование из каждого источника с разбивкой по годам, объем финансирования по факту и текущее состояние выполнения. Строительство или подготовку объекта выполняет генеральный подрядчик, в качестве государственного заказчика, как правило, выступает исполнительный орган государственной власти. Для любого из видов работ, выполняемых на объекте, фиксируются исполнители работ (услуг), параметры контракта. Работы, выполняемые на объектах, могут разбиваться на этапы.

При входе в подсистему интерактивной обработки информации по учету и контролю за мероприятиями, проводимыми на территории Астраханской области пользователь проходит аутентификацию и получает

доступ к определенным формам и данным в соответствии с привилегиями. После этого запускается главная форма (рис.2).

Для начала работы необходимо выбрать министерство или проект межведомственного характера. После этого становится доступен список соответствующих мероприятий, для каждого из которых можно последовательно вывести на экран и отредактировать объекты, параметры исполнителей работ, контрактов, этапы работ по каждому из объектов. Во всех формах предусматривается возможность получать информацию об исполнителях, ответственных за выполнение того или иного мероприятия.

Форма, предназначенная для просмотра и редактирования объектов, относящихся к категориям «Строительство и реконструкция» и «Реставрация памятников истории и культуры», позволяет отображать графические изображения объекта (исторический вид объекта, начальное состояние до реставрации /реконструкции, текущее состояние и проектное состояние) и текстовую информацию «Историческая справка об объекте» (рис. 3).

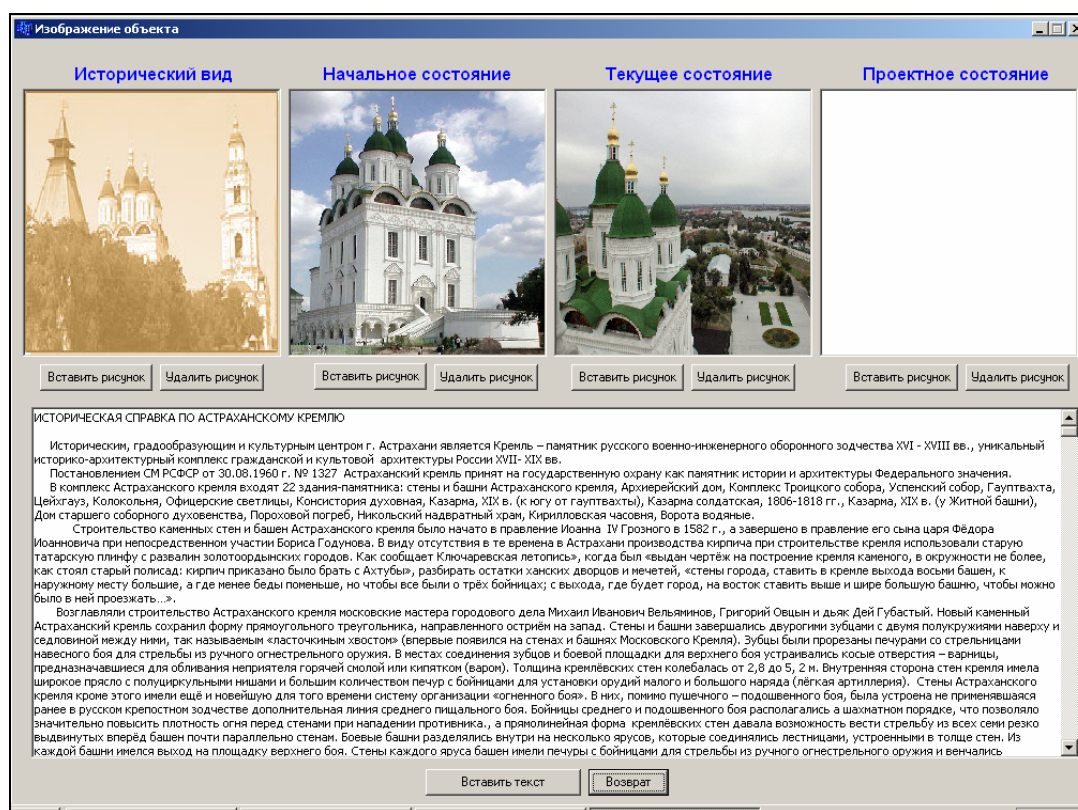


Рис.3 Форма изображений объекта и исторической справки

Основными функциями web-подсистемы являются:

- наличие системы управления контентом, предусматривающей авторизацию доступа с административным и клиентским интерфейсом, для управления содержанием и структурой сайта;
- масштабируемость структуры сайта, позволяющей без изменения структуры всего сайта детализировать структурные разделы, добавляя внутри них подразделы;
- организация поиска по контенту Интернет-сайта по одному или нескольким подряд идущим словам, неполному слову, либо с помощью интеграции с поисковыми системами (Яндекс, Google);
- осуществление обратной связи с пользователями: прием отзывов и запросов пользователей по e-mail через заполнение специальных форм сайта;

- наличие внутренней системы обработки статистики посещений, позволяющей получить информацию о наиболее часто посещаемых страницах, о времени нахождения пользователя на странице, о поисковых запросах по сайту;
- организация голосований и ведения их статистики.

Web-подсистема отображает информацию о мероприятиях как в целом, так и в соответствии с определенными критериями поиска. Кроме того, на сайте представлена географическая электронная карта города Астрахань (рис. 4) с нанесенными на нее активными точками – объектами мониторинга, предназначенная для получения информации об объекте.

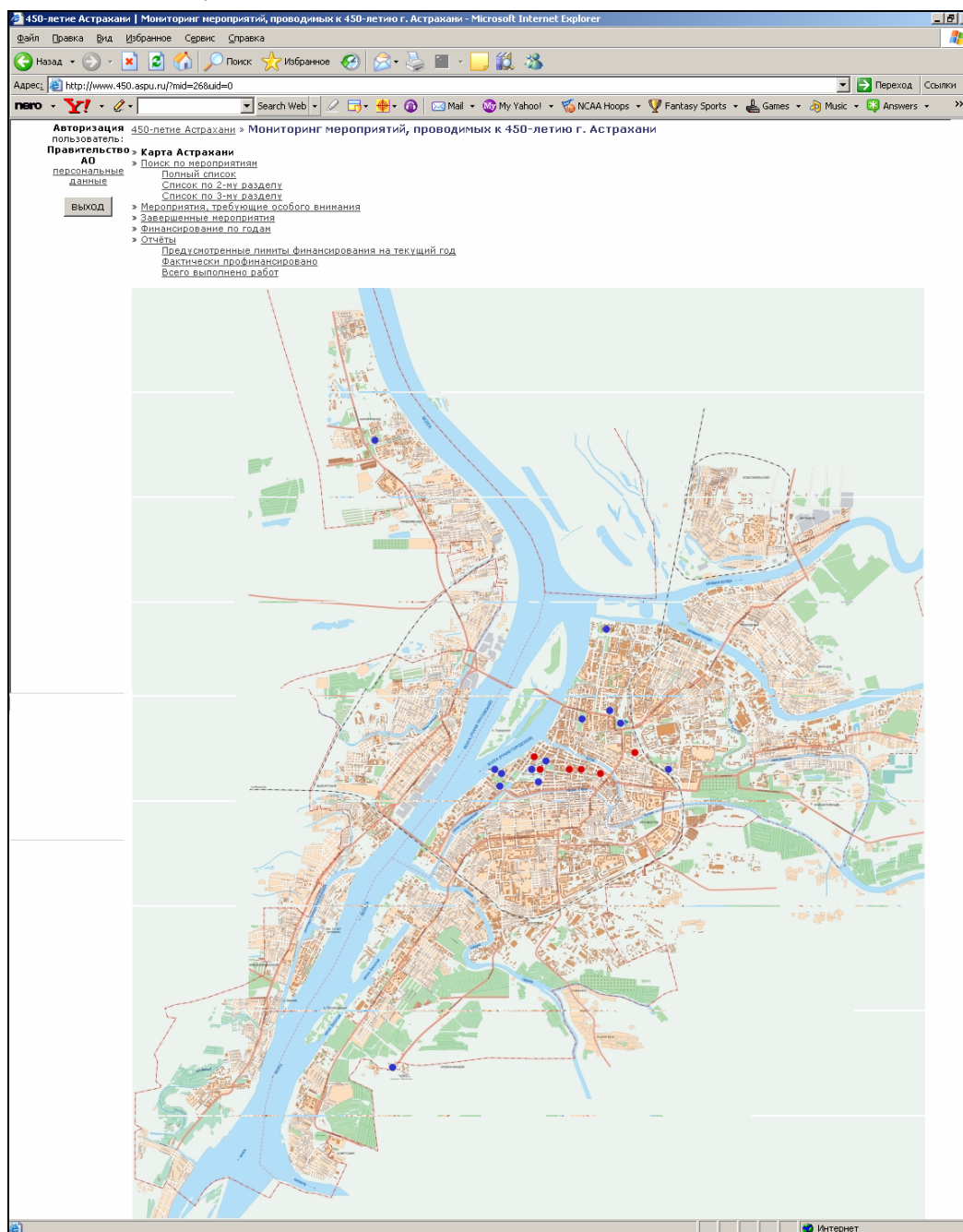


Рис. 4 Электронная карта г.Астрахань с нанесенными точками-объектами мониторинга

Каждый фрагмент карты может быть увеличен до размеров, позволяющих достаточно четко увидеть месторасположение того или иного объекта, на котором выполняются работы по строительству, реконструкции или реставрации (рис.5).

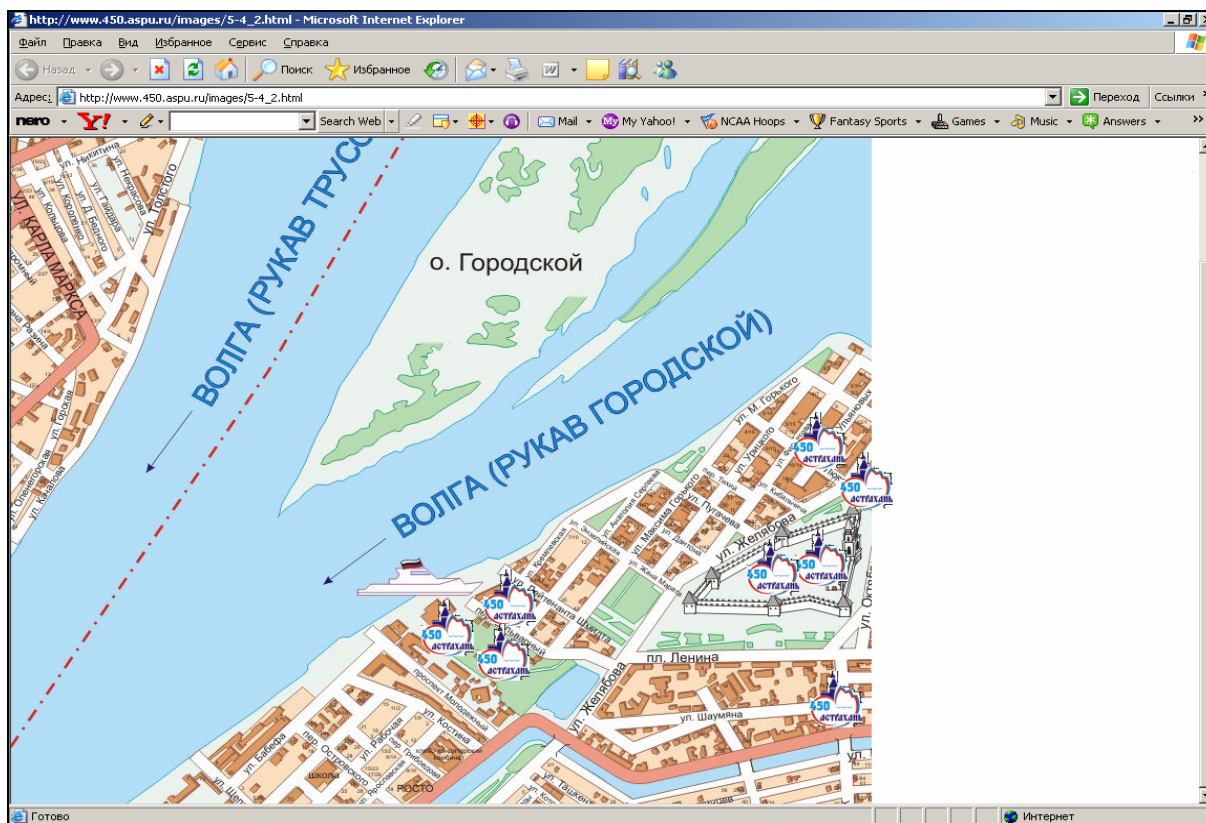


Рис. 5 Укрупненный фрагмент карты г.Астрахань.

PDA-версия системы представляет собой аналог web-версии, оптимизированный для просмотра на карманных персональных компьютерах с выходом в Интернет с помощью штатных web-браузеров.

Библиография

1. "Концепция использования информационных технологий в деятельности федеральных органов государственной власти до 2010 года". Распоряжение правительства Российской Федерации N 1244-р от 27 сентября 2004 года.

Информация об авторах

Петрова Ирина Юрьевна – проректор по инновациям и информатизации Астраханского государственного университета, д.т.н., проф.; Россия, г.Астрахань, ул.Татищева 20А; e-mail: petrova@aspu.ru

Зайцева Марина Анатольевна – руководитель агентства по печати и информационным коммуникациям Астраханской области, Россия, e-mail: infdep@astranet.ru

Щербинина Оксана Владимировна – начальник отдела АСУ Астраханского государственного университета, к.т.н., доцент, Россия, e-mail: oksana@aspu.ru

Эрман Евгений Анатольевич – начальник отдела Интернет-технологий Астраханского государственного университета, Россия, e-mail: evgeny.erman@aspu.ru

Тюрин Николай Александрович – ведущий программист отдела АСУ Астраханского государственного университета, Россия, e-mail: asu@aspu.ru

ОБ ОДНОМ ПОДХОДЕ К ПРОЕКТИРОВАНИЮ ИНТЕРФЕЙСА АВТОМАТИЗИРОВАННОЙ ДОКУМЕНТАЛЬНОЙ СИСТЕМЫ

Арсений Баканов

Аннотация: В статье рассматривается методика расчета, проектирования и оценки эргономичности интерфейса систем документооборота. Приведена оригинальная методика расчета и данные, полученные в ходе работ по проектированию интерфейса документальной системы.

Ключевые слова: проектирование интерфейса, дизайн, эргономика.

Введение

Возрастающая сложность задач, решаемых с использованием компьютерных систем, предъявляет высокие требования к их производительности и быстродействию. Однако для увеличения производительности и быстродействия компьютерных систем зачастую не достаточно просто увеличить тактовую частоту процессора, разрядность шины данных или емкость и количество HDD. Быстродействие и производительность компьютерных систем или систем “человек-компьютер” зависит не столько от производительности компьютера или быстродействия (скорости ввода информации/команд) пользователя, сколько от того, насколько будет понятно и полно представлена информация компьютером и насколько “понятными” для компьютера будут команды пользователя. Рассмотрим систему “человек – компьютер” представленную на рис. 1.

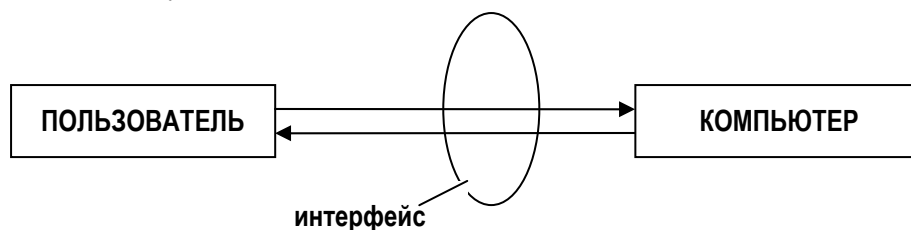


Рис. 1

Компьютер предоставляет информацию, на ее основе человек-пользователь принимает решение и осуществляет управляющее воздействие на компьютер, компьютер выполняет команды и предоставляет новую, актуальную информацию, на основе которой человек принимает решение и т.д. Производительность рассматриваемой системы будет зависеть от того насколько эргономично реализован интерфейс системы “человек – компьютер”. Рассмотрим вопросы оценки и проектирования интерфейса автоматизированных документальных систем. Существуют разнообразные методики, позволяющие количественно и качественно оценить эргономичность интерфейса.

Методики оценки эргономичности интерфейса

Для оценки эргономичности интерфейса используются следующие методы:

1. Метод экспертных/субъективных оценок

Оценка эргономичности интерфейса проводится группой экспертов, которые оценивают интерфейс, основываясь на своем, личном опыте работы. Поскольку эксперты оценивают один и тот же интерфейс по разному, основываясь каждый на своих субъективных критериях, целесообразно привлекать к оценке группу экспертов (от трех до пяти экспертов, так как при слишком большом количестве возрастает сложность анализа полученной информации).

2. Тестирование - сравнительный анализ

В тестировании эргономичности интерфейса принимают участие пользователи. Целью тестирования является выявление характерных ошибок пользователя при работе с интерфейсом программы. При проведении тестирования с участием пользователей целесообразно предлагать несколько вариантов интерфейса системы для того, чтобы пользователи выбрали наиболее эргономичный.

3. Количественные методы – многокритериальный анализ

Основными количественными показателями эргономичности интерфейса являются: скорость работы пользователя, скорость обучения, количество ошибок.

Традиционно для количественной оценки эргономичности интерфейса используют модель GOMS (the model of goals, objects, methods and selection rules) – «правила для целей, объектов, методов и выделения» [1]. Моделирование GOMS позволяет предсказать, какое количество времени потребуется пользователю для выполнения той или иной конкретной операции при использовании данной модели интерфейса.

Согласно модели GOMS время, требующееся для выполнения какой-либо поставленной задачи системой «пользователь-компьютер», является суммой всех временных интервалов, которые потребовались на выполнение последовательности действий, составляющих поставленную задачу.

Метод GOMS – дает хорошие количественные оценки вариантов интерфейсов пользователя. На практике разработчики пользуются расширенными моделями, такими как, например, CPM-GOMS (critical - path method GOMS), в которых учитываются дополнительные параметры.

Для оценки эргономичности документальных систем целесообразно использовать наряду с моделью GOMS совокупность коэффициентов с последующим проведением многокритериального анализа. Выше приведенные методики хорошо подходят для оценки уже существующих интерфейсов, однако их затруднительно использовать при проектировании интерфейсов.

Методика проектирования интерфейса

Рассмотрим методику проектирования интерфейса на примере автоматизированной системы Документационного Обеспечения Управления (ДОУ). Интерфейс рассматриваемой автоматизированной системы ДОУ представляет собой последовательность экранных форм с полями. Часть полей заполняется пользователем, часть полей заполняется автоматически. Каждое поле может принимать конечное число значений. Оценим информацию некоторой экранной формы. Будем измерять информацию экранной формы как уменьшение энтропии. То есть если рассмотреть экранную форму до заполнения (когда ее состояние было неопределено), то ее энтропия была $H(X_i)$, после заполнения всех полей (состояние полностью определилось), энтропия стала равной нулю. Следовательно, информацию, получаемую в результате заполнения всех полей, можно определить как

$$I_X = H(X_i) - 0 \quad \text{или} \quad I_X = H(X_i).$$

Все возможные состояния X_i экранной формы соответствующие им вероятности P_i представим в виде таблицы таб. 1.

Таблица 1

X_i	x_1	x_2	x_3	x_n
P_i	p_1	p_2	p_3	p_n

Согласно [1] энтропия определяется как

$$H(X_i) = - \sum_{i=1}^n p_i \log p_i,$$

где будем использовать логарифм по основанию два. Для рассматриваемой экранной формы определим следующие характеристики:

- пусть общее количество полей экранной формы согласно [2] не должно превышать $J = 7$;
- количество полей с предопределенным выбором (поля с прокруткой) не должно превышать $K = 5$;
- количество полей ввода (допускающих ввод не более пяти символов) не должно превышать $L = 4$;
- количество полей ввода с подсказкой (т.е. таких в которых система осуществляет ввод слова по двум первым буквам) не должно превышать $M = 3$;
- введем в рассмотрение некоторый коэффициент R_i характеризующий время затраченное программистом на разработку данного элемента интерфейса;
- каждому типу поля поставим в соответствие некоторый коэффициент t_{ij} характеризующий временные затраты оператора на заполнение данного поля. Тогда общее время на заполнение данной экранной формы

$$T_i(J) = \sum_{j=1}^J t_{ij}$$

При этих условиях, для данного конкретного случая можно сформулировать критерий производительности для интерфейса

$$H(X_i) \rightarrow \max \quad T_i(J) \rightarrow \min$$

Результаты анализа полей ввода сделанного в соответствие с критерием производительности приведены в таблице 2.

Таблица 2

Тип поля	Выбор из списка.	Ввод с подсказкой.	Ввод без подсказки.
Описание	До ввода информации все три состояния равновероятны, поэтому $H(X) = \log n$ [1]	В искомых 3 словах нас интересуют сочетания из 5 букв по 2. Число возможных равновероятных состояний 10.	Сочетания 5 букв из 32. Число возможных равновероятных состояний равно 32^5
Численное значение	$H(X_i) = \log 3 = 1,584$	$H(X_i) = \log 10 = 3,321$	$H(X_i) = 5 \log 32 = 25$
Коэффициент, характеризующий время, затраченное программистом на разработку элемента интерфейса	1	3	2
Коэффициент, характеризующий время, затраченное пользователем на заполнение поля	1	1,5	2

Выводы

В соответствие с данными, приведенными в таблице 2 можно сформулировать следующие рекомендации для проектирования интерфейса автоматизированной документальной системы - учитывая критерий производительности, временные затраты пользователя документальной системы на заполнение формы и временные затраты программиста-разработчика на создание интерфейса целесообразно использовать в рассматриваемом интерфейсе:

- полей с предопределенным выбором (поля с прокруткой) $K = 3$;
- количество полей ввода (допускающих ввод не более пяти символов) $L = 2$;
- количество полей ввода с подсказкой (т.е. таких в которых система осуществляет ввод слова по двум первым буквам) $M = 2$;

Таким образом, $J = 7$, т.е. общее количество полей экранной формы согласно [2] равно семи.

Литература

- Раскин Д. Интерфейс: новые направления в проектировании компьютерных систем. СПб., Символ-Плюс, 2004.
- Лурия А. Р. Лекции по общей психологии. СПб., Питер, 2006.
- Вентцель Е. С. Теория вероятностей. Москва. ФизМатЛит. 1962.
- Обознов А. А., Баканов А. С. Использование численных методов для оценки эргономичности интерфейса документальных систем. DCCN-2006. Moscow. 2006.

Сведения об авторе

Баканов Арсений Сергеевич – к.т.н., Институт психологии РАН, ул. Ярославская, д.13, Москва, Россия; e-mail: ilina@iitp.ru

РАЗРАБОТКА КОМБИНИРОВАННОГО МЕТОДА ДЛЯ ПРОЕКТИРОВАНИЯ СИСТЕМ ДОКУМЕНТООБОРОТА

Нина Баканова

Абстракт: В статье рассматриваются методы проектирования информационных систем для крупных организаций. Проводится сравнение структурного и объектно-ориентированного подхода. Для практической реализации автоматизированных систем документооборота предложен комбинированный метод разработки и анализа систем.

Ключевые слова: проектирование информационных систем, документооборот, делопроизводство.

Введение

Информатизация управленческой деятельности крупных организаций неразрывно связана с использованием автоматизированных систем документооборота. Рамочные функции систем документооборота определяются нормативами документационного обеспечения управления (ДООУ), как автоматизация технологических процессов прохождения, обработки и контроля исполнения документов. При этом в каждой конкретной организации эти системы, по сути, сопровождают основную деятельность и кроме перечисленных функций включают функции, связанные со спецификой отрасли управления [Фредерик Брукс, 2000]. Таким образом, проектировщики информационных систем вынуждены исследовать специфику документооборота в данной организации, особенности процедур принятия управленческих решений и целый ряд других факторов, влияющий на функциональные характеристики информационной системы.

Разработка проектов крупномасштабных информационных систем (ИС) сложной архитектуры невозможна без использования специальных методологий и автоматизированных инструментальных средств, обеспечивающих поддержку процессов проектирования, реализации и сопровождения функционирования ИС. В качестве таких средств выступают программные комплексы автоматизации проектирования или CASE-средства (Computer Aided Software Engineering), предназначенные для сокращения трудоемкости и сроков проектирования информационных систем за счет реализации специальных языков проектирования, создания коллективных средств разработки систем.

Постановка задачи

Для разработки проекта информационной системы документооборота крупной организации, требуется осуществить выбор методологии проектирования системы для наиболее эффективной работы по выявлению функциональных задач и структуры системы. При выборе методологии требуется учитывать, что работа будет выполняться как специалистами – аналитиками, так и специалистами прикладной области не знакомыми с методологиями и методами проектирования ИС.

В соответствии с системным подходом (**Structured Analysis and Design Technique SADT**) автоматизированная информационная система представляет собой совокупность взаимосвязанных объектов (элементов), функционирующих совместно для достижения общей цели [Д.А. Марка, 1997]. Существующие методологии проектирования реализуют представление взаимосвязанных объектов за счет использования различных графических нотаций для создания визуальной модели. При достаточной глубине проработки визуальная модель позволяет наглядно представить не только внутреннее строение системы, но и отразить основные особенности функционирования.

В методологии проектирования наиболее известны два основных подхода, принципиальное различие между которыми состоит в различных способах декомпозиции [С.В. Маклаков, 2003, А. М. Вендров, 2003]. В основу **функционально-модульного или структурного подхода** положен принцип функциональной декомпозиции, при котором модель системы описывается в терминах иерархии функций и передачи информации между отдельными функциональными элементами. При **объектно-ориентированном подходе** используется объектная декомпозиция. В этом случае, структура системы описывается в терминах объектов и связей между ними, а поведение системы описывается в терминах обмена сообщениями между объектами [А.М. Вендров, 2003, Грейди Буч, 2000]. Достоинством функциональных моделей является реализация принципа проектирования «сверху вниз», что соответствует традиционным представлениям иерархических функций в организационных системах управления [А.А. Беляев, 2000].

Недостатком структурного подхода считается достаточно сложный переход к проектированию структуры данных и программного обеспечения, необходимость использования других методов и, соответственно, пакетов автоматизации. Например, для представления моделей данных, разрабатываемой программой системы требуется использовать диаграммы сущность – связь (**Entity-Relationship Diagram – ERD**), для определения внешних ссылок и хранилищ данных требуется использовать и диаграммы потоков данных **Data Flow Diagram (DFD)**. Использование различных видов диаграмм, усложняет работу проектировщиков и соответственно увеличивает время на разработку проекта.

Более прогрессивной методологией является объектно-ориентированный подход (ООП), в основе которого лежит объектная модель, представляющая реальный мир в виде совокупности взаимодействующих объектов. Методология объектно-ориентированного подхода базируется на принципах абстрагирования, модульности, полиморфизма. Согласованность моделей проектируемой системы на всех стадиях разработки является требованием объектно-ориентированного подхода [Грейди Буч, 2000].

Для реализации объектного подхода в виде CASE-средств, разработан специальный язык проектирования – UML (Unified Modeling Language). Язык UML представляет собой мощное средство проектирования и предназначен, для работы высококвалифицированных специалистов - проектировщиков информационных систем [Крэг Ларман, 2001], но является достаточно трудным для нахождения «общего языка со специалистами прикладной области».

Вариант решения

Как показывает анализ *безнадежных проектов* [Крэг Ларман, 2001], основные ошибки проектирования возникают на этапах концептуального проекта при согласовании функций системы со специалистами прикладной области, определении ее границ и учета перспектив развития. Поэтому для достижения одной из главных целей внедрения CASE-технологий – снабжения всех участников проекта общим языком

«понимания», наиболее эффективным оказывается комбинированный подход, использующий как структурный, так и объектно-ориентированный метод.

При всех достоинствах ООП, наиболее критичным и трудоемким в данном подходе остается правильный выбор объектов и дальнейшая их разработка. Критериями выбора объектов в первую очередь являются принципы повторного использования и оптимизация построения программного комплекса системы. Эти принципы далеко не очевидны для специалистов предметной области. С другой стороны, для специалистов – аналитиков определенными трудностями представляет прикладная область, которая может иметь массу особенностей не заметных на первый взгляд. Структурные диаграммы более просты для понимания и позволяют проводить анализ необходимых функций системы.

Сопоставляя эти проблемы и на основе имеющегося опыта разработки больших систем [Н.Б. Баканова, 2003], было предложено на начальных этапах проектирования для общения со специалистами предметной области и скорейшего изучения особенностей создаваемой системы, использовать структурный подход, обеспечивающий: четкое определение границ системы, определение входных и выходных потоков информации (контекстные диаграммы); процесс проектирования функций «сверху вниз», что соответствует традиционным представлениям иерархической организации работ; возможность построения вариантов логических схем обработки документов; функциональную детализацию процедур обработки документов, которая позволяет выявить недостатки существующих процессов (бесполезные, неуправляемые и дублирующиеся работы).

В комбинированном методе предлагается проводить согласование двух типов диаграмм: функциональных диаграмм и диаграмм логического уровня ООП после функциональной декомпозиции выполненной до уровня логических операций. Это позволяет максимально использовать знания специалистов прикладной области для выявления всех необходимых функций и построения непротиворечивого проекта системы.

Процесс выделения объектов для ООП проводится уже после работы с функциональными диаграммами на основе разработанных спецификаций. Такой подход показал хорошие результаты при проектировании систем автоматизации документооборота крупных организаций, в которых специалистами предметной области являлись специалисты по делопроизводству, не знакомые с существующими технологиями проектирования.

Выводы

Для максимально эффективного использования знаний специалистов предметной области при построении больших информационных систем предлагается использовать комбинированный метод проектирования, в котором на начальных этапах рекомендуется использовать структурный подход. Следующим шагом проектирования должна быть подготовка спецификаций по всем предусмотренным работам структурного подхода для выявления объектов и перехода к процессу ООП.

Указанный комбинированный метод проектирования был использован для создания проектов систем документооборота для ряда крупных организаций, включая: Президиум РАН, Министерство транспорта РФ и другие крупные организации.

Литература

- [Марка Д. А., 1997] Марка Д. А., Мак Гоун К. Методология структурного анализа и проектирования. Учебное пособие. – М.: МетаТехнология, 1997.
- [С.В. Маклаков, 2003] С.В. Маклаков., Создание информационных систем с ALLFusion Modeling Suite. – М.: Диалог-МИФИ, 2003.
- [А.М. Вендров, 2003] А.М.Вендров., Проектирование программного обеспечения экономических информационных систем., М.: Финансы и статистика, 2003.
- [Грейди Буч, 2000] Грейди Буч, Джеймс Рамбо, Айвар Джекобсон. Язык UML. Руководство пользователя., М., ДМК, 2000. (с. 429)

- [А.А. Беляев, 2000] А.А. Беляев, Э.М. Коротков., Системология организаций., М.ИНФРА-М, 2000.
- [Крэг Ларман, 2001] Крэг Ларман. Применение UML и шаблонов проектирования., М., Изд. дом Вильямс, 2001. (с 489)
- [Фредерик Брукс, 2000] Фредерик Брукс. Мифический человек-месяц или как создаются программные системы. С-Петербург, Символ-Плюс, 2000. (с.298)
- [Н.Б. Баканова, 2003] Н.Б. Баканова., Проблемы внедрения систем документооборота в государственных организациях. – М., 2003г., Международная практическая конференция «Электронный документооборот и документационное обеспечение управления в бизнесе».

Сведения об авторе

Нина Борисовна Баканова - к.т.н., доцент, Институт проблем передачи информации РАН, Б. Каретный пер., Москва, 127994, Россия; e-mail: nina@iitp.ru

МЕТОД АНАЛИЗА ИЕРАРХИЙ КАК ИНСТРУМЕНТ ОБОСНОВАНИЯ БИЗНЕС-РЕШЕНИЙ

Михаил Плаксин

Аннотация: Метод анализа иерархий предложен Т. Саати около 30 лет назад для принятия решений в условиях неопределенности. Он может использоваться для решения задач управления, в том числе, задач прогнозирования и стратегического планирования. В данной статье предложен подход для применения этого метода в решении задач прогнозирования и стратегического планирования.

Keywords: метод анализа иерархий, принятие решений, прогнозирование, стратегическое планирование, парные сравнения, противоречивые критерии, неопределенность.

ACM Classification Keywords: H.4 Information Systems Applications: H.4.2 Types of Systems – Decision support.

Описания метода анализа иерархий

Большинство бизнес-систем можно отнести к сложным недетерминированным системам. Принятие управленческих решений в этих системах требует учета множества критериев, зачастую плохо формализованных и противоречащих друг другу. Кроме того, часто отсутствуют объективные показатели, по которым можно сопоставить альтернативные варианты для того, чтобы выбрать оптимальный. Для обоснования таких решений нужны специальные методы. Один из них – метод анализа иерархий (МАИ), разрабатываемый в последние три десятилетия американским ученым Т. Саати [Саати, 1991; Саати, 1993]. МАИ позволяет упорядочить работу лица, принимающего решение, и учесть достаточно сложную систему факторов, влияющих на выбор решения. Для отражения взаимосвязи этих факторов они организуются в виде иерархии, что и определило название метода.

Суть метода состоит в следующем. Пусть нам нужно решить некоторую проблему. Зафиксируем критерии, которым должно удовлетворять решение проблемы ($K_1, K_2, \dots, K_n$) и ее возможные решения ($R_1, R_2, \dots, R_m$). Задача состоит в том, чтобы из предложенных решений выбрать наилучшее. Согласно МАИ выбор будет производиться следующим способом (метод описывается предельно упрощенно).

1. Первый шаг: *построение матриц «парных сравнений» (МПС)*. Сначала мы должны попарно сопоставить друг с другом все критерии, сравнивая их важность для решения данной проблемы. Затем

для каждого из критериев мы выполняем попарное сравнение возможных решений, оценивая их важность *относительно данного критерия*. В результате получим одну матрицу $n \times n$ и m матриц $m \times m$.

Если для сравнения двух альтернатив (критериев или решений) нельзя использовать объективные показатели (вес, длина, стоимость и т.п.) будем применять следующую семнадцатизначную шкалу: 1 – равная важность альтернатив, 3 – слабое превосходство строки над столбцом, 5 – среднее превосходство строки над столбцом, 7 – сильное превосходство строки над столбцом, 9 – абсолютное превосходство строки над столбцом. $1/3$ – слабое отставание строки от столбца, $1/5$ – среднее отставание строки от столбца, $1/7$ – сильное отставание строки от столбца, $1/9$ – абсолютное отставание строки от столбца. 2, 4, 6, 8, $1/2$, $1/4$, $1/6$, $1/8$ – промежуточные значения.

2. Второй шаг: *синтез «локальных приоритетов»*. В каждой из матриц парных сравнений выполняем следующие действия: считаем среднее геометрическое каждой строки и нормализуем эти величины к единице, для чего делим каждое среднее геометрическое на сумму средних геометрических. Полученные нормализованные средние геометрические объявляем локальными приоритетами альтернатив, соответствующих строкам матриц. В результате получаем вектор локальных приоритетов для критериев и матрицу из локальных приоритетов, каждая строка которой соответствует одному возможному решению, а каждый столбец – одному критерию.

3. Третий шаг: *проверка согласованности матриц парных сравнений*.

Несо согласованность МПС выражается в нарушении транзитивности отношения превосходства (если $A > B$ и $B > C$, то должно быть $A > C$). Нарушение это может носить характер качественный ($A > B$, $B > C$, $C > A$; это возможно, поскольку альтернативы сравниваются попарно), а может – количественный. Количественная несо согласованность связана с тем, что мы не просто говорим о превосходстве той или иной альтернативы, но пытаемся оценить это превосходство по некоторой шкале (слабое/среднее/сильное). Отсюда возможна ситуация: $A >> B$, $B >> C$, $A > C$.

Дело осложняется тем, что при сравнении необходимо оперировать качественными понятиями (сильный-слабый), а в МПС мы записываем их числовое обозначение.

Причина несо согласованности матрицы заключается в следующем. Вообще говоря, содержимое МПС определяется n числами – «истинными весами» альтернатив. В идеале матрица должна состоять из отношений этих n чисел. Но поскольку истинных весов мы не знаем, мы вынуждены выбрать для матрицы $(n^2 - n)/2$ различных значений (размер матрицы – $n \times n$, но на главной диагонали стоят единицы, а левый нижний треугольник заполнен величинами, обратными по отношению к величинам из правого верхнего треугольника), которые будут служить некоторыми приближениями для отношений истинных весов. При построении матрицы парных сравнений обеспечивается ее обратная симметричность, но в матрице, построенной из отношений истинных весов, связи между элементами гораздо более многообразны, поскольку вес каждой альтернативы входит в $(2n - 1)$ элементов матрицы (все элементы одной строки и одного столбца).

В МАИ существует простая процедура, которая позволяет легко определить «оценку согласованности» МПС. Даются рекомендации о том, какую оценку можно считать хорошей, какую – приемлемой. К сожалению, в классических работах Т. Саати [Саати, 1991; Саати, 1993] ничего не говорится о том, как определить место несо согласованности МПС и устранить эту несо согласованность. Рекомендации по этому поводу можно найти в работах [Митягин, 2001; Плаксин, 2005].

4. Четвертый шаг: *синтез глобальных приоритетов*. Умножаем указанную матрицу локальных приоритетов, соответствующих решениям, на вектор-строку приоритетов, соответствующих критериям. Получим вектор-столбец приоритетов, соответствующих решениям.

5. Последний пятый шаг: *оценка согласованности всей иерархии*.

Описанный метод естественным образом распространяется на случай многоуровневой иерархии.

Использование метода анализа иерархий для решения задач прогнозирования

Одна из областей приложения МАИ – прогнозирование будущего состояния сложных систем. В МАИ задача прогнозирования понимается следующим образом. Есть некая система, состояние которой некоторым образом изменяется со временем. Существуют *акторы* (действующие лица), которые способны повлиять на развитие системы. Причем в зависимости от способа действия этих акторов система может перейти в то или иное *состояние*. Различные акторы по-разному относятся к разным состояниям системы. Какие-то из состояний для каждого актора более желательны, какие-то – менее. Каждый из акторов преследует свои цели и стремится воздействовать на систему так, чтобы она оказалась именно в том состоянии, которое наиболее выгодно именно этому актору. Акторы, вообще говоря, неравноправны. Степень их влияния на будущее системы различна. Задача в том, чтобы определить, в каком состоянии окажется система в результате суммарного воздействия всех акторов.

Для обозначения «способа действий» того или иного актора в МАИ используется термин «политика».

Отметим, что речь в данном случае идет о недетерминированных системах, поэтому оценка может быть только вероятностной.

Задача прогнозирования требует ответа на следующие вопросы:

- 1) Каким образом описать состояние исследуемой системы?
- 2) Каким образом описать изменение состояния системы в будущем?
- 3) Каким образом определить наиболее вероятный ход развития системы и то состояние, в котором система окажется?

МАИ отвечает на эти вопросы следующим образом.

Прежде всего, фиксируется некоторый *набор параметров*, с помощью которых будет описываться состояние системы. Для обозначения этих параметров в МАИ используется термин «переменные состояния». Параметры могут быть любыми, как количественными, так и качественными. Важно только одно: именно по изменению этих параметров мы будем оценивать изменение состояния системы.

Количество параметров может быть произвольно большим. Проблема в том, что с увеличением числа переменных состояния резко увеличивается трудоемкость анализа, что ведет не к увеличению, а к уменьшению точности результатов. Приходится искать некий компромисс между точностью модели и трудоемкостью оценки.

Значение *переменной состояния* призвано отразить изменения в состоянии исследуемой системы. Это легко сделать, если эти изменения могут быть измерены в некоторых объективных единицах (килограммах, метрах, рублях и пр.). Однако в большинстве случаев МАИ применяется именно тогда, когда объективные единицы измерения отсутствуют. Для решения этой проблемы МАИ предлагает следующий путь. Исходим из того, что нас будет интересовать не абсолютная величина того или иного параметра, а ее изменение по сравнению с текущим состоянием, не то, какого уровня достигнет та или иная характеристика, а то, в какую сторону и насколько она изменится по сравнению с сегодняшним днем. Соответственно, значение переменной состояния будет отражать не абсолютную величину некоторой системной характеристики, а ее изменение относительно базового уровня.

Для оценки изменения используется уже знакомая нам *нечеткая 17-значная шкала*. Только числовое выражение качественных изменений теперь несколько изменим. Это будут числа от -8 до $+8$, которые интерпретируются следующим образом: 0 – отсутствие изменения, ± 2 – слабое возрастание/убывание, ± 4 – среднее возрастание/убывание, ± 6 – сильное возрастание/убывание, ± 8 – максимально возможное возрастание/убывание. Нечетные числа интерпретируются как промежуточные состояния

Следующий шаг после выбора переменных состояния – *описание сценариев возможного развития событий*. Каждый сценарий задается как совокупность значений переменных состояния. Поскольку переменные состояния – это те параметры, отслеживая которые мы будем оценивать изменение

состояния исследуемой системы, а значение переменной состояния характеризует изменение некоторого параметра системы относительно базового уровня, каждый сценарий будет описывать совокупность изменений, которые произойдут в исследуемой системе. Каждый сценарий будет описывать некоторое возможное будущее состояние исследуемой системы, указывая по каким параметрам будущее состояние системы будет отличаться от нынешнего состояния и насколько сильным будет это отличие. Такие сценарии называются контрастными или исследовательскими.

Каждый такой сценарий имеет имя (краткое обозначение из одного двух слов), краткое описание (из одной фразы, выражающей суть сценария), «числовое» описание в виде совокупности значений переменных состояния, «словесное» описание в виде расшифровки этих значений, перевода их на естественный язык.

С помощью метода анализа иерархий *оценивается вероятность реализации каждого из контрастных сценариев*. Далее возможны два случая. Либо контрастные сценарии альтернативны и мы обязательно должны выбрать ровно один из них (например, выбор невестой одного из возможных женихов в моногамных странах). Таковым будет наиболее вероятный сценарий. Либо итоговое решение является неким компромиссом между рассматриваемыми альтернативами (например, государственный бюджет, принимаемый Госдумой с учетом влияния каждой из думских фракций). Для этого случая вводится понятие «обобщенного сценария». Обобщенный сценарий строится на основе контрастных с учетом вероятности осуществления каждого из них. В этом случае контрастные сценарии должны представлять собой некую полную (в некотором смысле) совокупность.

Для оценки вероятности осуществления различных сценариев строится *иерархия* из следующих уровней: *первичные факторы* (экономические, социальные, политические, технологические и т.п.); *акторы* (действующие лица); *цели акторов*; *политики* (возможные способы действия) *акторов*; *контрастные сценарии*; *обобщенный сценарий*. При необходимости количество уровней может быть увеличено или уменьшено.

Доминирование будет определяться ответами на следующие вопросы. Какой из первичных факторов оказывает большее влияние на будущее рассматриваемой системы? Кто из акторов оказывает большее влияние на тот или иной первичный фактор? Какая из целей важнее для того или иного актора? Какая из политик данного актора больше способствует достижению этим актором той или иной из своих целей? Осуществление какого из контрастных сценариев больше соответствует той или иной политике актора?

Рассчитанные по этой иерархии веса контрастных сценариев интерпретируются как вероятности их осуществления и используются либо непосредственно (для альтернативных сценариев), либо для синтеза обобщенного сценария.

Использование метода анализа иерархий для решения задач стратегического планирования

Описанное выше применение метода анализа иерархий для решения задач прогнозирования расширяется на задачи стратегического планирования. В МАИ стратегическое планирование понимается следующим образом.

В результате решения задачи прогнозирования мы оцениваем вероятность осуществления того или иного исследовательского сценария. В задаче стратегического планирования этот процесс называется *«первой прямой волной»*. Результат прямой волны – выбор наиболее вероятного сценария. Поскольку наиболее вероятный сценарий совсем не обязательно является наиболее желательным, возникает вопрос: а нельзя ли каким-либо способом увеличить вероятность того, что события будут развиваться в нужном нам направлении. Для поиска ответа на этот вопрос предназначена *«первая обратная волна»*. Она начинается с выбора наиболее желательного сценария и направлена на поиск таких политик, которые позволят увеличить вероятность осуществления именно этого сценария. Для этого строится *иерархия* из следующих уровней: *контрастные сценарии* (из них будет выбран наиболее желаемый); *проблемы*, мешающие осуществлению наиболее желательного для нас сценария; *акторы*, влияющие на решение

этих проблем (набор акторов может и не совпадать с набором акторов прямой волны); цели акторов; наши *политики*, которые призваны способствовать решению проблем, которые мешают осуществлению наиболее желательного для нас сценария. (К сожалению, МАИ не может дать никаких рекомендаций о том, какие это должны быть политики.) Результат обратной волны – новые политики, применение которых должно увеличить вероятность осуществления нужного нам сценария.

Обратим внимание на следующие детали. Прямая волна строилась «нейтрально», без выделения какого-либо актора. Все акторы были равноправны. Обратная волна строится с точки зрения какого-то актора. Появляются понятия «желательность сценария» (для кого?), «наши политики» (чьи?). И тем не менее, на обратной волне учитывается не только выбранный актор, но и другие действующие лица. Расчет прост. Если наши цели противоречат целям других акторов, другие акторы будут нам противодействовать. Поэтому желательно выбрать такие политики, которые будут наиболее приемлемы не только для нас, но и для других акторов.

Выбранные в результате обратной волны политики *добавляются в иерархию*, построенную для прямой волны, после чего проводится «*вторая прямая волна*». Ее цель – оценить, насколько политики, выбранные на обратной волне, приблизят желаемое для нас будущее. При необходимости можно провести вторую обратную волну, третью прямую и т.д. Существует способ оценки «сходимости» данного процесса.

Опыт применения МАИ, полученный магистрами кафедры математического обеспечения вычислительных систем Пермского госуниверситета, показывает, что этот метод может быть успешно использован для выбора стратегии развития предприятий самого разного рода.

Пути уменьшения трудоемкости метода анализа иерархий

Одним из главных недостатков метода анализа иерархий является его чрезвычайно высокая «экспертность». Она связана с двумя моментами: необходимостью большого числа парных сравнений и несогласованностью (противоречивостью) экспертных оценок.

В качестве иллюстрации потребности в большом числе парных сравнений рассмотрим вышеупомянутую иерархию для решения задач прогнозирования. Саати позволяет каждому элементу иерархии иметь до 15 потомков. Пусть их будет только 7. Тогда анализ иерархии потребует заполнить 113 МПС размером 7х7: одна матрица для сравнения факторов, 7 – для оценки акторов относительно каждого из факторов, 7 – для сопоставления целей каждого из семи акторов, 49 МПС – для сравнения политик акторов по отношению к их целям (у семи акторов целей 49). 49 МПС – для оценки контрастных сценариев относительно 49 политик, которые могут быть реализованы акторами.. При увеличении количества потомков до 10 количество МПС возрастает до 221, а размер матрицы – до 10х10.

Для снижения экспертной трудоемкости метода были предложены следующие механизмы:

- 1) построение МПС, согласованных «по построению»;
- 2) уменьшение числа сравнений за счет исключения из иерархии части элементов;
- 3) ускорение поиска рассогласованностей в матрице парных сравнений, выбор элементов, согласование которых даст максимальный эффект.

Для построения МПС, согласованных «по построению», разработана специальная «таблица сложения сравнений» (ТСС).

Применение ТСС базируется на следующем факте. Рассмотрим тройку альтернатив a_i, a_j, a_k с весами $p_i,$

p_j, p_k . Элементы МПС, соответствующие сравнению этих альтернатив друг с другом, будут равны $\frac{p_i}{p_j}, \frac{p_i}{p_k},$

$\frac{p_j}{p_k}$. (в числителе – вес «элемента-строки», в знаменателе – вес «элемента-столбца»). Тогда после

заполнения экспертом любых двух элементов МПС возникает желание третий вычислить автоматически. Такому автоматическому вычислению препятствует нечеткость шкалы сравнений. Традиционная таблица умножения в данном случае не годится. Например, если a_i имеет очень слабое превосходство над a_j , a_j – среднее превосходство над a_k , то как оценить отношение a_i и a_k ? Для ответов на подобные вопросы была разработана специальная таблица «сложения сравнений». Таблица составлена в терминах качественной шкалы и предлагает для каждой пары «складываемых сравнений» диапазон возможных вариантов (очень слабое превосходство + очень слабое превосходство = от очень слабого до среднего превосходства, слабое превосходство + среднее превосходство = от средне-сильного до сильного превосходства и т.д.). Качество ТСС было проверено при автоматической генерации около 600 тыс. иерархий. Все полученные с ее помощью МПС имели хорошую степень согласованности (до 10%). (По Саати степень согласованности до 10% считается хорошей, до 20% – терпимой, свыше 20% – неприемлемой.)

При ручном заполнении МПС таблица сложения сравнений может использоваться для оперативного контроля действия эксперта. Начиная со 2-й строки, вводимые экспертом элементы МПС могут с помощью ТСС проверяться на соответствие ранее введенным значениям. Таким образом можно выявить несогласованность МПС непосредственно во время построения. При этом необходимо иметь в виду два момента. Во-первых, несоответствие «нижних» строк «верхним» может означать неточность как «внизу», так и «наверху». Нельзя считать «верхние» элементы «более правильными» только потому, что они были записаны ранее «нижних». Во-вторых, значения из ТСС являются не более, чем рекомендацией, которую эксперт волен принимать или не принимать.

Еще одним свойством МПС, которое хочется использовать для уменьшения нагрузки на эксперта при заполнении матрицы, является ее обратная симметричность (элементы, симметричные относительно

главной диагонали, должны быть обратны: над диагональю – $\frac{p_i}{p_j}$, значит под диагональю – $\frac{p_j}{p_i}$). Кажется

естественным использовать эксперта для заполнения только одной половины матрицы, а вторую получить автоматически. Делать этого ни в коем случае нельзя! Как показывает опыт, при заполнении полной матрицы, называемые экспертом оценки далеко не всегда являются обратно симметричными. Более того, иногда оценка меняется не только количественно, но и качественно (при начальном сравнении $a_i > a_j$, при повторном – $a_j > a_i$). Поэтому обратную симметричность можно и должно использовать, но не для автоматизации заполнения МПС, а для контроля принимаемых экспертом решений.

Второй механизм – исключение из иерархии некоторых элементов – основывается на следующих рассуждениях. Качество вырабатываемых рекомендаций зависит (как минимум) от трех факторов: качества иерархии, квалификации эксперта и качества заполнения экспертом матриц парных сравнений. Увеличивая детальность иерархии, мы можем увеличить качество иерархии. Но при этом стремительно возрастает количество и размер МПС (пример был приведен выше). Соответственно, возрастает нагрузка на эксперта, их заполняющего. Для качественного выполнения работы требуется высококвалифицированный специалист. Время таких работников – дорого. Поэтому увеличение нагрузки ведет к сокращению внимания, уделяемого каждому конкретному вопросу, ведет к снижению качества заполнения матриц. В результате может оказаться, что увеличивая детализацию иерархии, мы не повышаем, а понижаем качество модели.

Возникает вопрос: существует ли способ определить, какая степень детализации иерархии является наилучшей? Ответ будет содержать в себе противоречие: и нет, и да. Нет – потому что это невозможно сделать статически, в момент построения иерархии. Да – потому что это оказывается можно сделать динамически во время анализа иерархии, поэтапного продвижения от ее корневой вершины к концевым.

В ходе построения иерархии мы не имеем еще никакой информации о значимости того или иного ее элемента. Поэтому в это время у нас нет никаких оснований принимать решение о том, следует ли включать тот или иной элемент в иерархию или без него можно обойтись. Поэтому чем более подробная

иерархия будет построена, тем лучше. Иное дело – динамика. Процесс анализа иерархии заключается в том, что мы двигаемся сверху вниз, заполняем матрицы парных сравнений и вычисляем сначала локальные, а потом и глобальные веса элементов. При этом элементы естественным образом кластеризуются (например, все акторы, все цели одного актора, все политики одного актора). Появляется возможность в каждом кластере отделить весомые, значимые элементы от незначимых. И незначимые – отбросить. Например, пусть первоначально мы включили в иерархию 10 акторов. После вычисления их глобальных весов выяснилось, что решение проблемы на 80% зависит только от четырех акторов (вес каждого – 15-25%), а суммарный вклад остальных шести акторов – 20% (вес каждого 3-4%). Возникает вопрос: стоит ли продолжать анализ этих шести акторов, если от них почти ничего не зависит?

Такое динамическое прореживание иерархии позволяет значительно сократить трудоемкость анализа. Без него анализ крупных иерархий становится попросту невозможным.

Прореживание иерархии приводит к вопросу о корректности этого действия, о том, какое влияние оно окажет на конечный результат. Как скажется исключение того или иного элемента на весах контрастных сценариев, т.е. на тех рекомендациях, которые будут выработаны в результате применения МАИ? В приведенном нами примере картина была достаточно ясна, разрыв между лидерами и аутсайдерами достаточно велик как по их индивидуальным весам (15-20% против 3-4), так и по суммарным (80% против 20). В практических задачах картина не всегда настолько контрастна. Отсюда – два вопроса, ответы на которые должны определять допустимость прореживания. Как скажется на конечном результате отбрасывание элемента с тем или иным весом? (Элементы какого веса еще можно отбросить, а какого – уже нельзя?) Как скажется на конечном результате отбрасывание группы элементов с тем или иным суммарным весом? (Каков может быть суммарный вес отбрасываемых элементов?)

К сожалению, точных ответов на эти вопросы пока не найдено. По полученным оценкам в случае отбрасывания акторов с суммарным весом в 10% вероятность смены лидера в списке контрастных сценариев составляет около 5%. При отбрасывании 20% эта вероятность возрастает до 10%. При отбрасывании 30% – до 15%. При отбрасывании 40% – до 20%.

Поскольку оценка влияния прореживания на конечный результат неточна, было предложено заменить прореживание более мягкой операцией – сверткой. Заключается она в том, что часть наименее значимых элементов кластера исключается из рассмотрения, но «не совсем». Вместо них в кластер включается новый элемент с именем «И другие» и с весом, равным суммарному весу исключенных элементов. Таким образом, исключенные элементы продолжают влиять на принятие решения. Но поскольку все они свернуты в один единственный элемент, их дальнейший анализ уже не требует трудозатрат. (Свернув акторов-аутсайдеров, мы избавляемся о необходимости анализировать их цели и политики.) В том случае, если элемент «Другие» должен появиться в МПС, во всех позициях строки и столбца матрицы, относящихся к этому элементу, ставится одно и то же «нейтральное» значение, не влияющее на веса остальных элементов ($1/n$, где n – количество сравниваемых альтернатив).

Как показывает опыт, замена прореживания на свертку всегда приводит к некоторому «сглаживанию» результатов. Разница между весами контрастных сценариев становится менее ощутимой. Примерно в 10% случаев прореживание и свертка приводят к существенно разным результатам: к тому, что у контрастных сценариев не просто меняются веса, а в списке контрастных сценариев происходит смена лидера.

Третий механизм сокращения экспертной емкости связан с устранением несогласованностей в МПС. Как уже было сказано выше в разделе, посвященном описанию метода, рекомендации по этому поводу приведены в работах [Митягин, 2001; Плаксин, 2005].

Библиографический список

- [Митягин, 2001] Митягин М.В., Плаксин М.А. Некоторые усовершенствования метода анализа иерархий // Математика программных систем: Межвуз. сб. науч. тр. / Перм. ун-т, Пермь, 2001. С. 56-66.
- [Плаксин, 2005] Плаксин М.А. Некоторые трудности применения метода анализа иерархий и пути их преодоления // Первая международная конференция «Системный анализ и информационные технологии» САИТ-2005 (12-16 сентября 2005 г., Переславль-Залесский, Россия): Труды конференции. В 2 т. Т. 1. М.: КомКнига, 2005. С. 291-297
- [Саати, 1991] Саати Т., Кернс К. Аналитическое планирование. Организация систем. М.: "Радио и связь", 1991. 224 с.
- [Саати, 1993] Саати Т. Принятие решений. Метод анализа иерархий. М.: "Радио и связь", 1993. 320 с.

Сведения об авторах

Михаил Плаксин – Пермский государственный университет, доцент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: mapl@list.ru.

ВЫЯВЛЕНИЕ НЕСООТВЕТСТВИИ В МОДЕЛИ ГЕНА ИНФОРМАЦИОННОЙ СИСТЕМЫ

Виктор Левыкин, Максим Евланов

Аннотация: В данной статье рассматриваются вопросы выявления несоответствий в модели гена информационной системы путем анализа особенности операций, осуществляемых над информацией в динамической мультистабильной информационной системе. Для анализа системы применяется возврат Ферми-Паста-Улама (ФПУ) – явление распределения энергии первоначального возмущения по высшим гармоникам с последующим сбором в спектр первоначального возмущения

Анализ причин возникновения противоречий в ходе исследования бизнес-процессов, разработки, адаптации, внедрения и сопровождения информационных систем (ИС) показал следующее. В большинстве случаев эти противоречия и причины, их вызвавшие, являются следствием изначального представления ИС как равновесной устойчивой системы. В связи с этим актуальной становится проблема разработки методологии, методов и моделей проектирования и эксплуатации ИС, в основе которых лежит представление ИС как динамической мультистабильной системы [1]. Основываясь на подобном представлении, авторы предложили концепцию информационного гена (ИГ). Под ИГ понимается в общем случае упорядоченная и в сильной степени сжатая последовательность знаний (правил) построения ИС в целом и ее отдельных элементов. По своей сути ИГ является метамоделью ИС - формализованным представлением, которое определяет синтаксис и семантику конкретных реализаций ИС и ее компонентов [2]. Хотя термин «метамодель» в этой трактовке приобрел более узкий смысл, суть его осталась той же, что и в традиционной теории систем, где существование метаописания и метамодели является и необходимым, и конструктивным [3]. Наиболее приемлемым способом организации метамодели представляется реализация некоторого механизма, обеспечивающего раздельную организацию работы механизма логического вывода и механизма интерпретации результатов вывода [4].

Использование концепции ИГ позволяет рассматривать процесс проектирования ИС как синергетическое взаимодействие отдельных элементов системы. Данное взаимодействие позволит решить проблему выявления и устранения противоречий, возникающих из-за неоднозначного восприятия предметной области участниками проекта.

Для выявления этих противоречий авторы предлагают проанализировать особенности операций, осуществляемых над информацией в динамической мультистабильной ИС. В результате этого анализа можно сделать вывод, что нарушение устойчивых состояний ИС является следствием выполнения исключительно операций генерации новой информации [1], поскольку только такие операции формируют новые устойчивые состояния. Поэтому выявление и устранение противоречий следует проводить по результатам анализа генерации и распространения новой информации на категорно-топологической модели ИГ, объектами которой являются атрибуты и структуры атрибутов информационного пространства ИС, а морфизмами – операции, осуществляемые над этими атрибутами и структурами атрибутов.

Выполнение конкретной операции генерации новой информации может привести ИС к одному из следующих состояний:

- 1) операция генерации новой информации осуществляется одним элементом ИГ, а сгенерированная информация распространяется в ИС в соответствии со структурными особенностями ИГ;
- 2) операция генерации новой информации осуществляется одним элементом ИГ, однако на пути распространения сгенерированной информации наблюдаются обрывы маршрутов распространения информации;
- 3) одна и та же операция генерации новой информации осуществляется несколькими элементами ИГ, а распространение сгенерированной информации в соответствии со структурными особенностями ИГ приводит к противоречивости результатов выполнения этих операций в ряде элементов ИС.

Первое состояние является следствием отсутствия противоречий в процессе проектирования ИС. Второе состояние является следствием возникновения противоречий, вызванных обособленностью представлений участниками проекта предметной области и проектируемой ИС. Третье состояние является следствием противоречий, вызванных конфликтами относительно данных между участниками проекта.

Для выявления этих состояний предлагается представить категорно-топологическую модель ИГ как колебательную систему. При этом колебания в данной системе порождаются выполнением операций генерации новой информации и распространяются в ИС в результате выполнений операций рецепции и хранения информации. В таких системах вне зависимости от их природы наблюдается возврат Ферми-Паста-Улама (ФПУ) – явление распределения энергии первоначального возмущения по высшим гармоникам с последующим сбором в спектр первоначального возмущения. Особо интересным свойством возврата ФПУ оказалось наличие «памяти» в его спектре по отношению к начальным условиям его активных мод [5].

Применительно к ИГ наблюдение возврата ФПУ означает, что значения отдельных атрибутов, образующих локальные структуры или даже целые документы, после однократного генерирования многократно используются по отдельности или в самых различных комбинациях, а при необходимости могут быть возвращены в состояние, наблюдавшееся при выполнении операции генерации новой информации. Иными словами, если для изучаемого ИГ наблюдается возврат ФПУ, это означает, что информация, которая вводится в соответствующую систему, может быть разделена на отдельные

составляющие и впоследствии воспроизведена без ошибок. При этом процесс функционирования ИС представляет собой последовательность операций по обработке информации. В том случае, если для изучаемого ИГ возврат ФПУ не наблюдается или наблюдается с искажениями, это означает, что одна и та же информация дублируется в ИС и в ходе распространения взаимно поглощается или искажается. При этом появляется возможность заранее выявить области информационного пространства такой системы, в которых будут наблюдаться конфликты относительно данных, и предупредить эти конфликты.

В этом случае рекомендуется осуществлять анализ разрабатываемой ИС путем изучения возможности наблюдения возврата ФПУ двумя основными способами. Первый способ наиболее целесообразно применять на предпроектных стадиях и в процессе проектирования типовых ИС. Он подразумевает частичное знание генерируемых локальных структур и документов, а также невозможность установления точных характеристик операций генерации новой информации в типовой ИС. В этом случае изучается только факт выполнения операции генерации новой информации без учета особенностей ее реализации на конкретном объекте функционирования. Второй способ наиболее целесообразно применять в процессе эксплуатации типовой ИС. Тогда становится возможным уточнение модели ИГ и, в частности, модели операций генерации новой информации применительно к особенностям бизнес-процессов конкретного объекта функционирования. Проведение анализа в этом случае позволит выявить основные проблемы, затрудняющие эксплуатацию ИС, установить конкретные направления работ по модернизации системы и ее адаптации к изменению бизнес-процессов.

Предлагаемый подход позволяет выявлять и устранять противоречия в ИГ ИС уже на ранних стадиях ее проектирования, а именно, в процессе формирования и анализа требований к ИС. Кроме того, данный подход может использоваться для решения проблемы интеграции отдельных проектных решений в единую целостную ИС, а также для проведения мониторинга эксплуатируемых на конкретных предприятиях ИС с целью выявления наиболее типичных особенностей и тенденций изменения их бизнес-процессов.

Литература

1. Чернавский Д.С. Синергетика и информация (динамическая теория информации). - М.: Едиториал УРСС, 2004. - 288 с.
2. Фаулер М., Скотт К. UML в кратком изложении. Применение стандартного языка объектного моделирования. - М.: Мир, 1999. - 191 с.
3. Месарович М., Такахара Я. Общая теория систем: математические основы. - М.: Мир, 1978. - 312 с.
4. Лачинов В.М., Поляков А.О. Информодинамика, или Путь к Миру открытых систем. - СПб.: Издательство СПбГТУ, 1999.
5. Гаряев П.П. Волновой геном. - М.: Общественная польза, 1994. - 280 с.

Информация об авторах

Проф., д.т.н. **Виктор Макарович Левыкин**, зав. кафедрой Информационно-Управляющих Систем - Харьковский национальный университет радиоэлектроники, г. Харьков, Украина

к.т.н. **Максим Викторович Евланов** – доцент кафедры Информационно-Управляющих Систем - Харьковский национальный университет радиоэлектроники, г. Харьков, Украина

СИНЕРГИЧЕСКИЕ АСПЕКТЫ ПРОЦЕССА ПОИСКА РАВНОВЕСНОЙ ЦЕНЫ ОБЪЕКТА НЕДВИЖИМОСТИ

Марина Чичужко, Валерий Тазетдинов

Аннотация: В статье предложен метод определения равновесной цены объекта недвижимости, который базируется на использовании нейросетевых технологий и позволяет активизировать процесс функционирования рынка недвижимости.

Ключевые слова: Объект недвижимости, равновесная цена, нейросетевые технологии.

Введение

Современное состояние рынка недвижимости является уменьшенной копией состояния украинской экономики в целом. Постепенно заканчивается период быстрого накопления капитала с его сверхдоходами и происходит переход к обществу, в котором значительная часть финансовых потоков производится сферой информационных технологий. Уменьшение нормы прибыли вызывает интерес к изучению рынка, его субъектов, структуры и механизмов функционирования, причин и следствий принятия решений. Растет необходимость в использовании информационно-аналитических методов прогнозирования ситуации и предвидения будущих сценариев развития, которые нужны для потенциальных инвесторов, органов государственного управления с целью принятия стратегических решений по развитию бизнеса и усовершенствованию механизмов функционирования рынка.

Внедрение современных технологий в информационно-аналитическую инфраструктуру обслуживания клиентов предусматривает, в первую очередь, оптимизацию поведения субъектов рынка недвижимости и обеспечения достижения ими своих целей. Методы аналитической обработки данных позволят получить информацию о максимально возможной стоимости объекта для продавца, обеспечить осуществление покупки за минимальную цену, увеличить прибыль за счет сокращения времени операций продажи, увеличения количества клиентов и улучшения качества обслуживания.

Разработка и применение информационно-аналитического обеспечения процессов на рынке недвижимости предопределено необходимостью внесения порядка в его структурную и элементную базу, которая даст возможность проведения целеустремленной государственной политики.

Постановка задачи

При анализе номенклатуры объектов недвижимости установлено, что, чаще всего, значения их характеристик имеют равномерное или нормальное распределение. В качестве базового объекта выберем среднестатистический объект недвижимости (ОН) со средними характеристиками – внутренними параметрами. Известно, что на стоимость ОН влияют и внешние факторы. Коррекция цены ОН в зависимости от особенностей местоположения, города, региона, результатов сравнительного анализа со столичными ценами и макроэкономической ситуаций в стране является важной задачей в процессе принятия решений. Для ее решения предложим такой метод:

Для каждого из внешних факторов $Y_1, Y_2, \dots, Y_q$ построим функции принадлежности $\mu(Y_1), \mu(Y_2), \dots, \mu(Y_q)$, причем $\mu(Y_i) \in [0, 1]$, $i = \overline{1, q}$. Нормируем значение внутренних факторов.

Выполним структурную идентификацию функции оценки ОН как зависимости

$$Z' = F(X_1, \dots, X_n, Y_1, \dots, Y_q, a_0, \dots, a_n, b_1, \dots, b_q). \quad (1)$$

В случае предположения о линейном характере вышеприведенной зависимости будем искать функцию

$$Z' = a_0 + \sum_{i=1}^n a_i X_i + \sum_{j=1}^q b_j \mu(Y_j) \quad (2)$$

при выполнении ограничений на коэффициенты:

$$\sum_{i=0}^n a_i + \sum_{j=1}^q b_j = 1. \quad (3)$$

Заметим, что в таком случае значения факторов путем нормирования необходимо отобразить в отрезок $[0,1]$.

Алгоритм поиска равновесной цены

Предложенный метод не является единственно возможным [Марков, 1989]. Достаточно точные результаты получаем и при использовании корректирующих коэффициентов, но эмпирический характер их определения ставит больше вопросов, чем приносит пользы.

Интенсификация строительства нового жилья определяет и задачу формирования цены, которая устраивала бы и продавцов и покупателей. Эффективным рынок недвижимости (РН) является в том случае, когда наблюдается равенство предложения и спроса [Витлинский, 2003]. Будем считать, что Z_t – цена ОН в момент времени t , D_t , S_t количество однотипных ОН, купленных и проданных за время t . Тогда строительные организации для адекватного реагирования на конъюнктуру рынка должны учитывать следующие математические соотношения:

$$S_t = f(Z_{t-1}), D_t = g(Z_t), \quad (4)$$

$$\lim_{t \rightarrow \infty} f(Z_{t-1}) = \lim_{t \rightarrow \infty} g(Z_t), \lim_{t \rightarrow \infty} Z_t = Z_q, \quad (5)$$

где $f(Z)$ – монотонно растущая и $g(Z)$ – монотонно убывающая функции, Z_q – равновесная цена. Имея статистический материал, функции $f(Z)$ и $g(Z)$ можно идентифицировать с помощью нейросети (НС) и, как следствие, найти точку равновесной цены. Отличия ОН требуют введения определенных корректирующих процедур, которые базируются на построении трапециевидной функции принадлежности (рис. 2). С абсолютной мерой уверенности ($\mu_q = 1$) считаем, что цена ОН в диапазоне $(x_q - \varepsilon, x_q + \varepsilon)$ позволит продать этот объект за время t . Чем меньшим будет это значение, тем меньшим является время реализации ОН. В интервалах $(x_q - \delta, x_q - \varepsilon)$ ОН также будет продан, но время этой процедуры будет значительно больше.

Конструктивно задачу определения равновесной цены ОН будем решать в такой последовательности. Используем две взаимосвязанные НС с алгоритмом обратного распространения ошибки и принцип стохастической релаксации. В качестве выходных данных имеем таблицу данных предложения (ИБ₁) и таблицу данных спроса (ИБ₂), которые содержат только значимые факторы в числовом формате. Одним из основных полей таблиц есть фактор времени, значение которого является количеством периодов, в течение которых осуществлялся мониторинг цены на ОН. Тогда нейронные сети осуществляют идентификацию превращений:

$$F_d : (X_u, t) \rightarrow Z_d, \quad (6)$$

$$F_s : (X_u, t) \rightarrow Z_s, \quad (7)$$

где $X_u \subset X$ – вектор наиболее значимых компонентов вектора X .

Без ограничения общности допускаем, что $X_u = (X_1, X_2, \dots, X_k), k < n$. Доказано [Горбань, 1996], [Колмогоров, 1957], что уже двухслойная НС аппроксимирует любую непрерывную функцию с заданной точностью. К особенностям аппроксимации зависимостей (6) и (7) вернемся ниже, а пока будем считать, что обученные НС осуществляют идентификацию функций F_s и F_d .

Поиск равновесной цены объективно приемлемой для продавца и покупателя осуществляется по следующему алгоритму:

Шаг 1. Пусть существует v основных факторов, значения которых определены заинтересованным субъектом рынка недвижимости, то есть $X_i = D_i, i = \overline{1, v}$. Значения других $k - v$ факторов для него являются неосновными.

Шаг 2. Сделав анализ записей в базе данных, для которых $X_i = D_i, i = \overline{1, v}$ определить минимальные и максимальные значения неосновных факторов, то есть $X_j^{\min} = D_j^{\min}$ и $j = \overline{v+1, k}$.

Шаг 3. Подать на вход обученных нейронных сетей векторы значений $(D_1, D_2, \dots, D_v, D_{v+1}^{\min}, D_{v+2}^{\min}, \dots, D_k^{\min})$ и $(D_1, D_2, \dots, D_v, D_{v+1}^{\max}, D_{v+2}^{\max}, \dots, D_k^{\max})$ и определить минимальные $Z_d^{\min}, Z_s^{\min}$ и максимальные $Z_d^{\max}, Z_s^{\max}$ цены для покупателя и продавца.

Шаг 4. Если интервалы $[Z_d^{\min}, Z_d^{\max}]$ и $[Z_s^{\min}, Z_s^{\max}]$ непересекаются, то определение равновесной цены или невозможно, или требует дополнительных исследований.

Шаг 5. Если интервалы пересекаются, то мера их пересечения, которую можно описать треугольной или трапециевидной функцией принадлежности, указывает на интервал, на котором находится равновесная цена и задача решена.

Если решение задачи определения равновесной цены является важным для сбалансированного функционирования рынка недвижимости, то определение влияния на цену ОН внутренних и внешних факторов является главным информативным фактором для строительных организаций при выборе оптимальной структуры строительства жилья и указывает на эффективное решение проблемы с его продаж. Эта задача декомпозируется на подзадачи согласно с количеством факторов, которые желательно учитывать (однофакторная или многофакторная) и согласно предположенному виду зависимости (линейная или нелинейная):

$$Z = f_1(X) \text{ (однофакторная),} \quad (8)$$

$$Z = f_2(X_{i1}, X_{i2}, \dots, X_{ip}) \text{ (многофакторная),} \quad (9)$$

$$Z = a_0 + \sum_{j=1}^p X_{ij} \text{ (линейная),} \quad (10)$$

$$Z = f_3(X_{i1}, X_{i2}, \dots, X_{ip}) \text{ (нелинейная).} \quad (11)$$

Если зависимость является линейной (10), то для определения абсолютного коэффициента чувствительности (влияния на цену объекта недвижимости изменения значения одного из факторов) достаточно ее продифференцировать по нужной переменной. В другом случае, который и определяется НС, коэффициент чувствительности не является постоянной величиной. Для его вычисления необходимо весь интервал значений фактора X_b разбить на одинаковые интервалы $\{x_p^0, x_p^1, \dots, x_p^h\}$ так, что $i = \overline{1, h}$.

Учитывая это, строим график изменения чувствительности исходной характеристики к изменению экзогенного фактора в зависимости от значений последнего как график зависимости $\Delta u = G(i)$. При меньшем значении δ зависимость будет точнее.

Поскольку значения экзогенных факторов и эндогенной характеристики являются нормируемыми, то процедуру определения чувствительности можно использовать и для нахождения значимых факторов. Как уже отмечено выше, количество факторов, которые учитываются при определении цены ОН, равно приблизительно тридцати. Идентификация функций спроса и предложения происходит при присутствии значительных шумовых эффектов, вызванных субъективностью суждений субъектов рынка недвижимости. Основным же принципом предварительной обработки данных является снижение существующей избыточности информации. Результатом его использования является повышение информативности начальных данных. Для определения факторов, которые несущественно влияют на

эндогенную характеристику, предлагаем использовать процедуру, в которой не используются никакие дополнительные предположения о виду зависимости между фактором X и ценой ОН Z .

Как предложено в [Ежов, 1998], формирование пространства признаков с учетом реальной значимости факторов осуществляется постепенно. На первом шаге определяют фактор с наибольшей индивидуальной значимостью, которая равняется

$$k_1 = \arg \max_{n_1} \{I(X_{n_1}, Z)\}, \quad (12)$$

где $I(X, Z)$ – количество информации, которая указывает на предсказуемость характеристики Z при известных значениях X и определяется в случае нелинейных зависимостей по технологии “box-counting” [Ежов, 1998]. Далее определяют второй, наиболее значимый фактор в паре с уже избранным:

$$k_2 = \arg \max_{n_2} \{I(X_{n_1} X_{n_2}, Y)\} \quad (13)$$

и процедуру продолжают к определению последовательности с заданным количеством значимых факторов.

Аналитическими методами обеспечить реализацию такой процедуры невозможно. В то же время ее достаточно просто выполнить с помощью НС. Для этого необходимо протабулировать значение коэффициентов чувствительности для каждого фактора на каждом промежутке (см. вышеприведенную процедуру). Упрощая задачу, найдем средние значения коэффициентов чувствительности эндогенной характеристики к изменению каждого фактора. Пусть для фактора среднее значение коэффициента чувствительности является наибольшим, что будет свидетельствовать о его значительной информативности для определения эндогенной характеристики. С помощью НС с одним входом и одним выходом, который функционирует по алгоритму обратного распространения ошибки найдем зависимость $Z_1 = f_1(X_{n_1})$. Далее рассчитаем отклонение:

$$Z - Z_1 = Z - f_1(X_{n_1}) = \Delta_1 = F_2(X_{i_1}, X_{i_2}, \dots, X_{i_{k-1}}), \quad i_j \neq n_1 \quad (14)$$

На следующем шаге формируем НС с вектором входов $(X_{i_1}, X_{i_2}, \dots, X_{i_{k-1}})$ и одним выходом, учим ее и определяем наиболее значимый фактор по значению коэффициента чувствительности Δ_1 к изменению указанных факторов. Процедуру рекуррентно повторяем до определения полной иерархии факторов по информативности. Практически вычисления прекращаются при определении заранее указанного количества значимых факторов.

Процесс определения общих тенденций ценовой политики почти тождественен получению решения предыдущей задачи. Главное отличие заключается в том, что необходимо получить зависимость цены от времени, в самом простом случае, и цены от времени и других факторов. Решение такой задачи позволит прогнозировать поведение субъектов рынка недвижимости и при необходимости осуществлять регуляцию и коррекцию политики строительных организаций.

Если считать заданным число кластеров, эталоны $x_e^i, i = \overline{1, K}$ в каждом кластере $Q_1, Q_2, \dots, Q_K$ и критериальные функции, то автоматически является определенными и классы ОН. Если же в результате предыдущего анализа установлено, что наибольший спрос имеют ОН с фиксированным значением фактора, то тогда необходимо решать задачу классификации, которая формулируется так:

$$\text{найти } \min_{x_e} \sum_{i=1}^K d_i \quad (15)$$

где d_i – расстояние между ОН i -го класса и соответствующим эталоном, причем

$$d_i = \left(\sum_{j=1}^{m_K} (x_j - x_e^i)^2 \right)^{\frac{1}{2}}, \quad (16)$$

x_e^i - координата эталона в i -м классе.

Получение ее решения и будет означать классификацию ОН (рис. 1).

Успешное функционирование РН зависит и от нахождения решения задачи идентификации зависимостей

$$\begin{aligned} D &= D(X, Z) = f(X_1, X_2, \dots, X_n, Z_d), \\ S &= S(D, Z) = g(X_1, X_2, \dots, X_n, Z_s) \end{aligned} \quad (17)$$

и поиску точки равновесной цены Z_e (рис. 2) [Витлинский, 2003], которая определяется равенством

$$f(X_1, X_2, \dots, X_n, Z_e) = g(X_1, X_2, \dots, X_n, Z_e) \quad (18)$$

при выполнении условий, что такая цена существует и где C_i – константа, D – количество однотипных ОН, купленных на РН, S – количество ОН, которые предлагаются на РН.

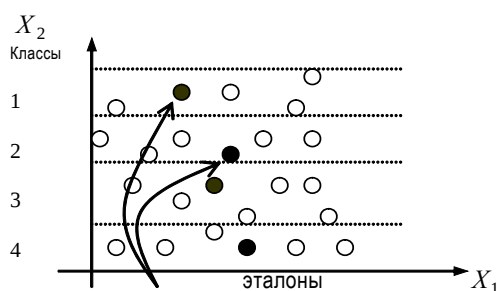


Рис. 1. Классификация с эталонами

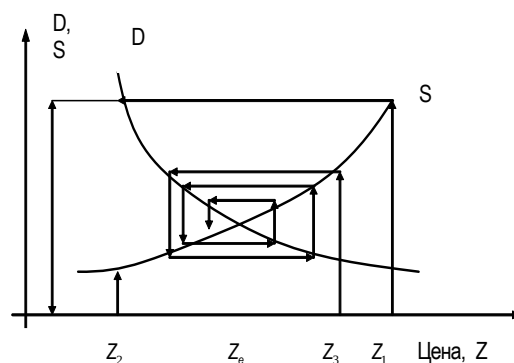


Рис.2. Процесс формирования равновесной цены

Заключение

Таким образом, использование разработанной нейросетевой технологии анализа и прогнозирования процессов на рынке недвижимости расширяет научно-методическую базу исследования рынка недвижимости, а ее достоверность подтверждается апробацией в агентствах недвижимости.

Библиография

- [Марков, 1989] А. Марков, Л. Расстригин. Синергетика: проблемы моделирования сложных систем и эволюционных процессов // Известия АН Латвийской. ССР. – Рига. – 1989. – № 6. – С. 134-136.
- [Витлинский, 2003] В.В. Витлинский. Моделирование экономики. – Киев: КНЕУ, 2003. – 408 с.
- [Горбань, 1996] А.Н. Горбань, Д.А. Россиев. Нейронные сети на персональном компьютере. – Новосибирск: Наука, 1996. – 256 с.
- [Колмогоров, 1957] А.Н. Колмогоров. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного // Докл. АН СССР, 1957. – Т. 114. – № 5. – С. 679-681.
- [Ежов, 1998] А.А. Ежов, С.А. Шумский. Нейрокомпьютинг и его применения в экономике и бизнесе. – Москва: МИФИ, 1998. – 224 с.

Информация об авторах

Марина Чичужко – Черкасский государственный технологический университет; бул. Шевченко, 460, Черкассы, Украина; e-mail: marina-puh@rambler.ru

Валерий Тазетдинов – Черкасский государственный технологический университет; бул. Шевченко, 460, Черкассы, Украина.

Summer School

CONCEPTS AND GOOD PRACTICES IN REPORTING, BUSINESS INTELLIGENCE AND DATAWAREHOUSING

Koen Vanhoof

Abstract: The lecture is aimed to explain the reason for different business intelligence technologies, to show the added value and to give best practices of datawarehousing; olap – multidimensional databases; data mining; integration.

Concepts and good practices in reporting, business intelligence and datawarehousing

Prof. Dr. Koen Vanhoof
Limburgs Universitair Centrum
koen.vanhoof@uhasselt.be

1.

Outline

Explain the reason for different business intelligence technologies, show the added value and give best practices

- Datawarehousing
- Olap – multidimensional databases
- Data mining
- Integration

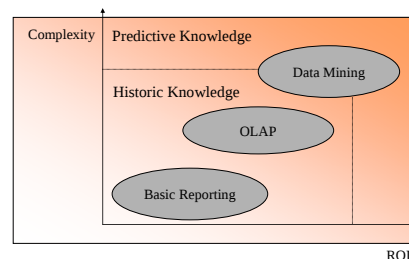
2.

Business Intelligence : Facts

- If companies are interested in analysing their customers, they are confronted with the following findings:
 - Acquiring far more expensive than retaining
 - Re-attracting customers far more expensive than keeping them satisfied.
 - Far easier selling a product to an existing customer.
 - Part of the customers vastly more profitable.
- This leads to a new method in treating its customers:
 - Moving from single sales models to continuing selling relationships with the customer, based on knowledge about this customer.

3.

Business Intelligence : Evolution



4.

Reporting - BI platform

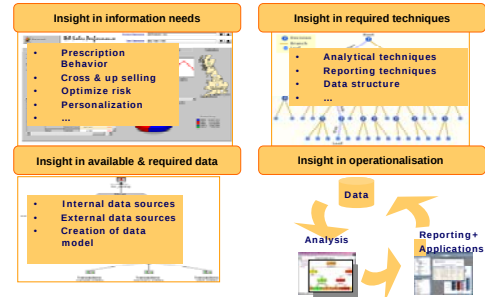
- Gather correct data in a structured way
- Availability of analysis techniques: Statistics, data mining, Time Series Analysis, linear programming,...
- Availability of reporting techniques: OLAP, Reporting Portal, A...

Data
Ware house



5.

Where is the Intelligence ?



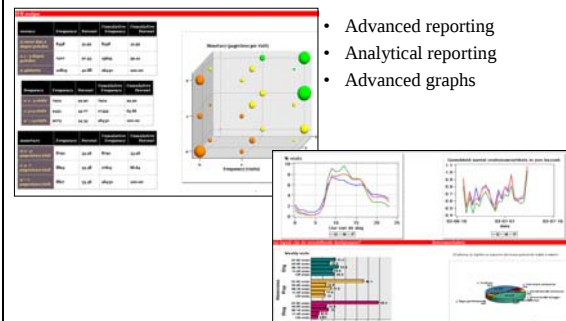
6.

From : Basic Statistics



7.

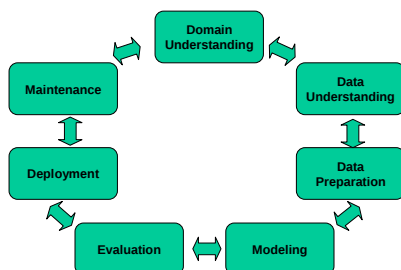
To : delivering the right insights



- Advanced reporting
- Analytical reporting
- Advanced graphs

8.

BI : What to do?



9.

Business Intelligence : Basic Questions

- What ? quantity, profit, ..
- When? time dimension
- How? sales channel
- Where? geographical
- Who? client

10.

Business Intelligence : Analytical questions

- Why (not) ?
- What if ?
- Prediction ?
- New initiatives (cross selling, up selling)?
- Which profiles

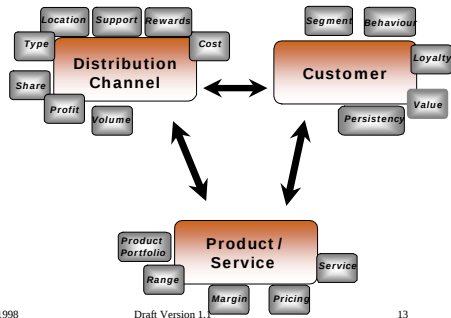
11.

Business Intelligence : Control questions

- Conform budget
- On time
- Quality of Service
- On target

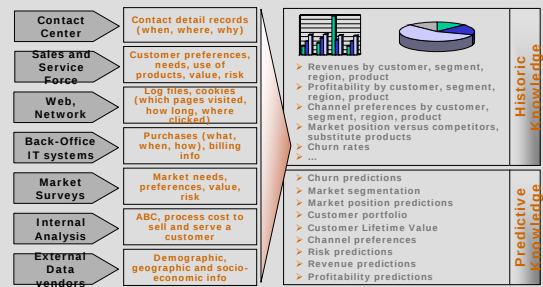
12.

BI requirement : Integration



13.

BI : Integration



14.

Data Warehousing

- However, operational systems were never designed to support such business activities.
- Businesses typically have numerous operational systems with overlapping and sometimes contradictory definitions.
- **Definition :** A subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process (Inmon, 1993).

15.

Subject-oriented Integrated Data

- Warehouse is organized around major subjects of the enterprise (e.g. customers, products, sales) rather than major application areas (e.g. customer invoicing, stock control, product sales).
- The integrated data source must be made consistent to present a unified view of the data to the users.

16.

Time-variant Non-volatile Data

- Data in the warehouse is only accurate and valid at some point in time or over some time interval.
- Data in the warehouse is not updated in real-time but is refreshed from operational systems on a regular basis.
- New data is always added as a supplement to the database, rather than a replacement

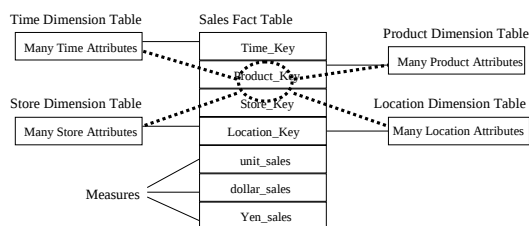
17.

Modeling Data Warehouses

- Modeling data warehouses: dimensions & measurement
 - *Star schema*: A single object in the middle connected to a number of objects (dimension tables) radially.
 - *Snowflake schema*: the dimensional hierarchy is represented explicitly by normalizing the dimension tables.
- Storage of selected summary tables:
 - *Independent summary table* storing pre-aggregated data, e.g., total sales by product by year.
 - *Encoding aggregated tuples* in the same fact table

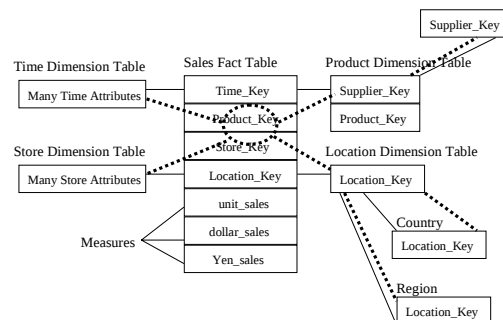
18.

Example of Star Schema



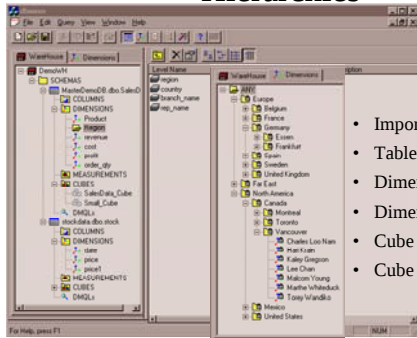
19.

Example of a Snowflake Schema



20.

View of Warehouses and Hierarchies



- Importing data
- Table Browsing
- Dimension creation
- Dimension browsing
- Cube building
- Cube browsing

21.

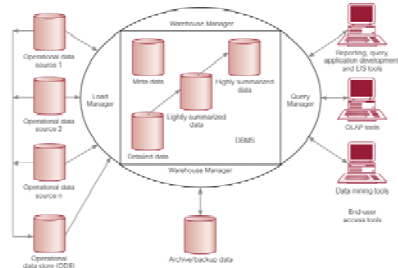
Comparison of OLTP Systems and Data Warehousing

Table 30.1 Comparison of OLTP systems and data warehousing systems.

OLTP systems	Data warehousing systems
Holds current data	Holds historical data
Stores detailed data	Stores detailed, lightly, and highly summarized data
Data is dynamic	Data is largely static
Repetitive processing	Ad hoc, unstructured, and heuristic processing
High level of transaction throughput	Medium to low level of transaction throughput
Predictable pattern of usage	Unpredictable pattern of usage
Transaction-driven	Analysis driven
Application-oriented	Subject-oriented
Supports day-to-day decisions	Supports strategic decisions
Serves large number of clerical/operational users	Serves relatively low number of managerial users

22.

Typical Architecture of a Data Warehouse



23.

Metadata

- Used for a variety of purposes
 - Extraction and loading processes - metadata is used to map data sources to a common view of information within the warehouse.
 - Warehouse management process - metadata is used to automate the production of summary tables.
 - Query management process - metadata is used to direct a query to the most appropriate data source.

24.

Metadata

- Copy management tools use metadata to understand the mapping rules to apply in order to convert the source data into a common form.
- End-user access tools use metadata to understand how to build a query.
- The management of metadata within the data warehouse is a very complex task that should not be underestimated.

25.

Data Warehouse Information Flows



26.

Outline

Explain the reason for different business intelligence technologies, show the added value and give best practices

- Datawarehousing
- **Olap – multidimensional databases**
- Data mining
- Integration

27.

Why Multi-Dimensional Databases?

- No single "best" data structure for all applications within an enterprise
- From econometric research conducted at MIT in the 1960s, the multidimensional database has matured into the database engine of choice for data analysis applications
- Inherent ability to integrate and analyze large volumes of enterprise data
- Offers a good conceptual fit with the way end-users visualize business data
 - Most business people already think about their businesses in multidimensional terms
 - Managers tend to ask questions about product sales in different markets over specific time periods

28.

What is a Multi-Dimensional Database?

A **multidimensional database (MDD)** is a computer software system designed to allow for the efficient and convenient storage and retrieval of large volumes of data that is (1) intimately related and (2) stored, viewed and analyzed from different **perspectives**. These **perspectives** are called **dimensions**.

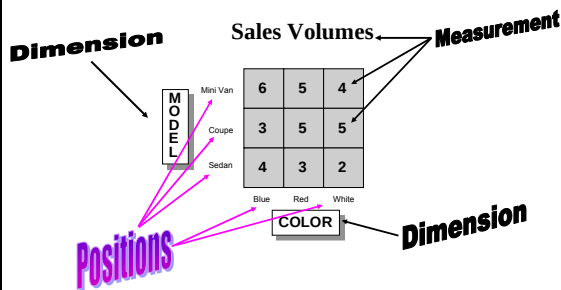
29.

Contrasting Relational and Multi-Dimensional Models: An Example

SALES VOLUMES FOR GLEASON DEALERSHIP

MODEL	COLOR	SALES VOLUME
MINI VAN	BLUE	6
MINI VAN	RED	5
MINI VAN	WHITE	4
SPORTS COUPE	BLUE	3
SPORTS COUPE	RED	5
SPORTS COUPE	WHITE	5
SEDAN	BLUE	4
SEDAN	RED	3
SEDAN	WHITE	2

30.



31.

Differences between MDD and Relational Databases

- Multidimensional array structure represents a *higher level of organization* than the relational table
- Perspectives are *embedded directly* into the structure in the multidimensional model
 - All possible combinations of perspectives containing a specific attribute (the color BLUE, for example) line up along the dimension position for that attribute.
- Perspectives are placed in *fields* in the relational model - tells us nothing about field *contents*.

32.

Differences Between MDD and Relational Databases

- MDD makes data browsing and manipulation intuitive to the end-user
- Any* data manipulation action possible with a MDD is also possible using relational technology
- Substantial *cognitive* advantages in query formulation
- Substantial *computational* performance advantages in query processing when using MDD

33.

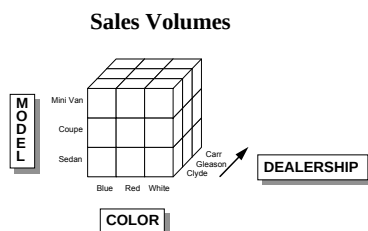
Contrasting Relational Model and MDD- Example 2

SALES VOLUMES FOR ALL DEALERSHIPS

MODEL	COLOR	DEALERSHIP	VOLUME
MINI VAN	BLUE	CLYDE	6
MINI VAN	BLUE	GLEASON	6
MINI VAN	BLUE	CARR	2
MINI VAN	RED	CLYDE	3
MINI VAN	RED	GLEASON	5
MINI VAN	RED	CARR	5
MINI VAN	WHITE	CLYDE	2
MINI VAN	WHITE	GLEASON	4
MINI VAN	WHITE	CARR	3
SPORTS COUPE	BLUE	CLYDE	2
SPORTS COUPE	BLUE	GLEASON	3
SPORTS COUPE	BLUE	CARR	2
SPORTS COUPE	RED	CLYDE	7
SPORTS COUPE	RED	GLEASON	5
SPORTS COUPE	RED	CARR	2
SPORTS COUPE	WHITE	CLYDE	4
SPORTS COUPE	WHITE	GLEASON	5
SPORTS COUPE	WHITE	CARR	1
SEDAN	BLUE	CLYDE	6
SEDAN	BLUE	GLEASON	4
SEDAN	BLUE	CARR	2
SEDAN	RED	CLYDE	1
SEDAN	RED	GLEASON	3
SEDAN	RED	CARR	4
SEDAN	WHITE	CLYDE	2
SEDAN	WHITE	GLEASON	2
SEDAN	WHITE	CARR	3

34.

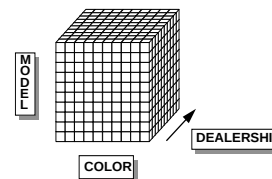
Multidimensional Representation



35.

Viewing Data - An Example

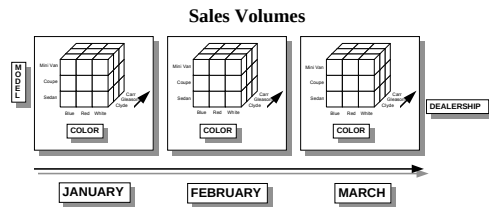
Sales Volumes



- Assume that each dimension has 10 positions, as shown in the cube above
- How many records would be there in a relational table?
- Implications for viewing data from an end-user standpoint?

36.

Adding Dimensions- An Example



37.

When is MDD (In)appropriate?

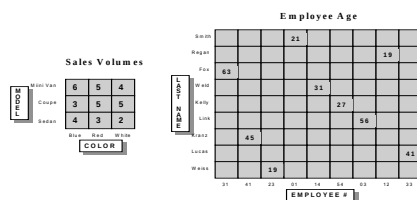
First, consider situation 1

PERSONNEL		
LASTNAME	EMPLOYEE#	EMPLOYEE AGE
SMITH	01	21
RELAN	12	19
ROK	31	63
VELD	14	31
KELLY	54	27
LINK	03	56
KRANZ	41	45
ILLUS	33	41
WEES	23	19

38.

When is MDD (In)appropriate?

MDD Structures for the Situations



Note the sparse between the two MDD representations

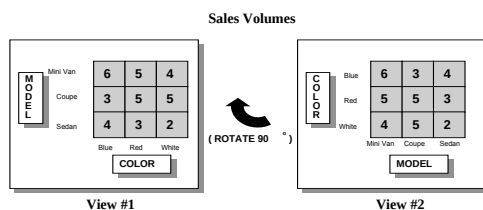
39.

When is MDD (In)appropriate?

- Our sales volume dataset has a great number of meaningful interrelationships
- Interrelationships more meaningful than individual data elements themselves.
- *The greater the number of inherent interrelationships between the elements of a dataset, the more likely it is that a study of those interrelationships will yield business information of value to the company.*
- Highly interrelated dataset types be placed in a multidimensional data structure for greatest ease of access and analysis

40.

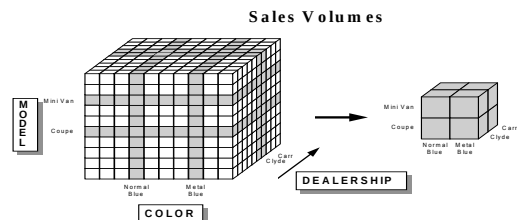
MDD Features - Rotation



- Also referred to as "data slicing."
- Each rotation yields a different slice or two dimensional table of data.

41.

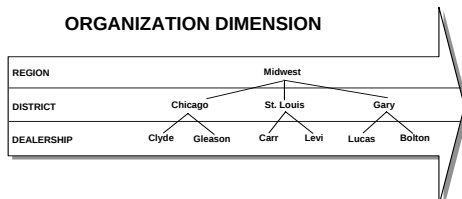
MDD Features - Ranging



- The end user selects the desired positions along each dimension.
- Also referred to as "data dicing."
- The data is scoped down to a subset grouping

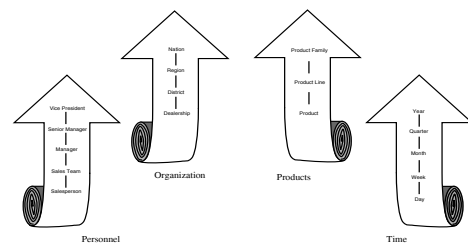
42.

MDD Features - Roll-Ups & Drill Downs



- The figure presents a definition of a *hierarchy* within the organization dimension.
- Aggregations perceived as being part of the same dimension.
- Moving up and moving down levels in a hierarchy is referred to as "roll-up" and "drill-down."

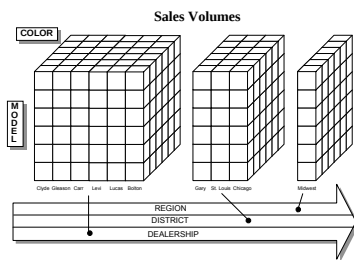
43.



Rolling-up and Drilling-down Through Multiple Dimensions. In the previous graphic, we saw that users can roll-up and drill-down through a single dimension, ORGANIZATION. Well designed multidimensional databases also allow users to roll-up and drill down through multiple dimensions concurrently. Thus, in this example, an end-user could hold the positions MANAGER, DISTRICT and PRODUCT constant, while a drilling-down or rolling-up through sales figures over the TIME dimension.

44.

MDD Features: *Drill-Down Through a Dimension*



45.

MDD Features: Multidimensional Computations

- Well equipped to handle demanding mathematical functions.
- Can treat arrays like cells in spreadsheets. For example, in a budget analysis situation, one can divide the ACTUAL array by the BUDGET array to compute the VARIANCE array.
- Applications based on multidimensional database technology typically have one dimension defined as a "business measurements" dimension.
- Integrates computational tools very tightly with the database structure.

46.

The *Time* Dimension

- TIME as a predefined hierarchy for rolling-up and drilling-down across days, weeks, months, years and special periods, such as fiscal years.
 - Eliminates the effort required to build sophisticated hierarchies every time a database is set up.
 - Extra performance advantages

47.

What is MDX ?

- MDX = Multi Dimensional Expressions
- A syntax for modeling and querying an OLAP database
- Part of the OLE DB for OLAP Spec
- Supported by multiple providers (OLAP Services, TM1, SAS, WhiteLight, SAP...)
- It is the key for all advanced analytical capabilities of OLAP Services

48.

MDX Basics

- MDX allows easy navigation in the multi dimensional space.
- It "understands" the MD concepts of cube, dimension, level, member and cell.
- It is used for
 - queries - full statements (SELECT...FROM)
 - business modeling - defining calculated members.

49.

MDX Queries

- SELECT sets on axes (N axes)
- FROM a source cube
- WHERE the data is sliced by some coordinates

50.

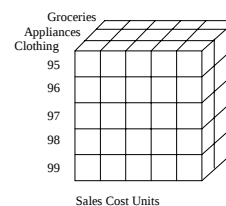
Sample Cube

- Dimensions :
 - Measures : Sales, Cost, Units
 - Time : Year, Quarter, Month
 - Geography : All, Region, Country
 - Products : All, Product Group, Product Name
 - Customers : All, Customer Group, Customer Name

51.

Every Cell has a name...

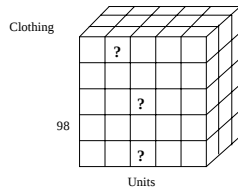
(Products.Clothing, Measures.Units, Time.98)
 (Products.Clothing, Measures.Sales, Time.97)
 (Products.Groceries, Measures.Cost, Year.95)



52.

Every Cell still has a name...

(Products.Clothing, Measures.Units, Time.98.PrevMember)
 (Products.Clothing, Measures.Units, Time.98.NextMember)
 (Products.Clothing, Measures.Units.PrevMember, Time.98.Lag(3))
 OR
 (Products.Clothing, Measures.Units.PrevMember, Time.98.Lead(-3))



53.

Calculated Members

- Calculated members are the real power behind OLAP.
- They provide the mean to define complex business logic.
- MDX is the syntax for calculated members
 - Statements - tools, query generators, developers
 - Calculated members - DBAs and power users

54.

Sales Growth

With member Measures.Growth as
 '(Time.currentmember,Measures.Sales)-
 (Time.currentmember.prevmember,Measures.Sales)'

Groceries				
Appliances				
Clothing				
95	15	13	5	15
96	17	13	6	2
97	22	11	7	5
98	29	9	9	7
99	30	10	8	1

Time.CurrentMember
 Time.99
 Time.CurrentMember.PrevMember
 Time.98

Sales Cost Units Growth

55.

How did sales in this periode compare with sales in the previous period?

		Sales	Result
1997	Q1	790	
		120	
	Jan	30	
	Feb	40	10
	Mar	50	10
	Q2	200	80
		65	15
	Apr	45	-20
	May	90	45
	Jun	185	-15
	Q3	55	-45
		60	5
	Jul	70	10
	Aug	285	100
	Q4	80	10
		100	20
	Oct	105	5
	Nov		
	Dec		

56.

How did sales in this period compare with sales in the previous period?

		Sales	Result
Q2		200	80
	Apr	65	15
	May	45	-20
	Jun	90	45

Time.CurrentMember, Measures.Current.Member
 Time.CurrentMember, Measures.Sales
 Time.CurrentMember.PrevMember, Measures.Sales

57.

Axis and Slicer Dimensions

- Axis Dimension : User wants to see the data against a lot of members in this dimension.
- Slicer dimension : User wants to choose a particular member in this dimension and see data against members in the axis dimensions.

58.

Set Expressions

- Simple enumeration : (USA, FRANCE, Paris)
- Geography.MEMBERS = set of all members in Geography dimension.
 - <level>.MEMBERS,<hierarchy>.MEMBERS, etc.
- Controlling member order :
 - HIERARCHIZE(<set>) puts each member below its parent. Siblings occur in "natural" order.
 - ORDER(<set>,<num-exp>) sorts set based on value of <num-exp>

* Can either retain or break hierarchy

59.

There are hundreds of MDX functions

. HIERARCHIZE	. EXTRACT	. MAX
. TOPCOUNT	. GENERATE	. VAR
. TOPPERCENT	. PERIODSTODATE	. STDDEV
. TOPSUM	. LASTPERIODS	. RANK
. BOTTOMCOUNT	. YTD	. AGGREGATE
. BOTTOMSUM	. QTD	. COVARIANCE
. FILTER	. MTD	. CORRELATION
. UNION	. WTD	. LINREGSLOPE
. DISTINCT	. SUM	. LINREGINTERCEPT
. INTERSECT	. COUNT	. LINREGVARIANCE
. EXCEPT	. AVG	. LINREGR2
. CROSSJOIN	. MEDIAN	. LINREGPOINT
	. MIN	

60.

Analytical Problems

- Explosion of dimensions
- Appropriate aggregation level
- Counting, adding , ... allowed
- Comparing totals, periods ...
- Explosion of reports

61.

Interpreting Mean

Scenario	Q1	Q2	Q3	Q4	Mean	Forecast
A	4	5	3	0	3	2
B	2	4	2	4	3	2
C	3	3	3	3	3	2
D	5	5	1	1	3	2
E	0	3	4	5	3	2

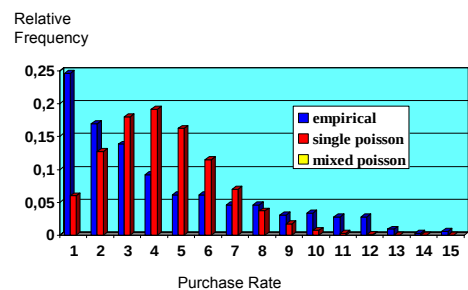
62.

Interpreting Counts

Client		A	B
Q1	Visits	18	16
	Coffee	5	2
	Coffee filter	2	1

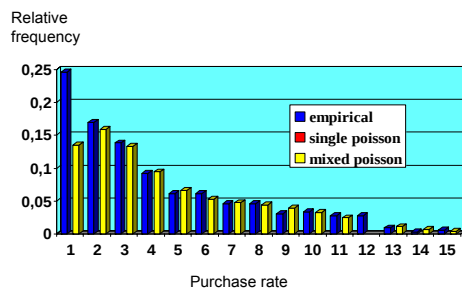
63.

Empirical and Estimated Distributions of Candy Purchases



64.

Empirical and Estimated Distributions of Candy Purchases



65.

Candy Results

Segment	%	Purchase rate
1	15.6	0.0
2	13.4	1.1
3	47.4	3.2
4	12.8	7.5
5	3.0	11.1
6	7.9	13.6

66.

Solutions

- Different measures
- Segmentations
- Mining for Interactions
 - supervised : prediction/classification
 - unsupervised : a priori

67.

Technical problems

- Timing
- Quality of data
- Performance / Storage / Maintenance
- Consistency

68.

Contrasting Relational Model and MD Model

Criteria	Relational Model	Multidimensional Databases
Focus	Data integrity of each piece of data	Facilitate exploration of interrelationships between dimensions
Organization structure	One-dimensional array	Multi-dimensional arrays
Perspectives	Embedded in fields	Embedded <i>directly</i> in MDDB structure
Computational power for Query processing	Joining tables often required; computationally expensive	Structure designed for OLAP; computationally cheap
Cognitive issues in querying data	Cumbersome	Intuitive
Query Languages	SQL or SQL front-ends, such as QBE	Point-and-click emphasis No standardized language
Management of Time Dimension	Not well suited	Well suited

69.

Benefits of MDD

- Cognitive Advantages for the User
- Ease of Data Presentation and Navigation
 - Obtaining the same views in a relational world requires the end user to either write complex SQL queries or use an SQL generator against the relational database to convert the table outputs into a more intuitive format.
- Ease of Maintenance
 - Because data is stored in the same way as it is viewed (i.e. according to its fundamental attributes), no additional overhead is required to translate user queries into requests for data
- Performance
 - Multidimensional databases achieve performance levels that are difficult to match in a relational environment.

70.

Outline

Explain the reason for different business intelligence technologies, show the added value and give best practices

- Datawarehousing
- Olap – multidimensional databases
- Data mining
- Integration

71.

Why Data Mining

- Limitation of traditional database querying:
 - Most queries of interest to data owners are difficult to state in a query language
 - “find me all records indicating fraud” => “tell me the characteristics of fraud” (Summarisation)
 - “find me who likely to buy product X” (classification problem)
 - “find all records that are similar to records in table X” (clustering problem)
 - Ability to support analysis and decision making using traditional (SQL) queries become infeasible (*query formulation problem*).

72.

The Query Formulation Problem

Consider the query :

What kinds of weather condition are suitable for playing tennis ?

- It is not solvable via query optimisation
- Has not received much attention in the database field or in traditional statistical approaches
- These problems are of *inductive* features: learning from data rather than search from data
- Natural solution is via train-by-example approach to construct *inductive models* as the answers

73.

What is Data Mining

Data Mining: a non-trivial data analysis process for identifying valid, useful and understandable patterns from databases.

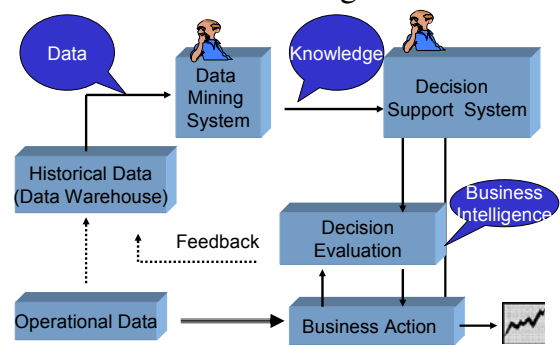
74.

What is Data Mining

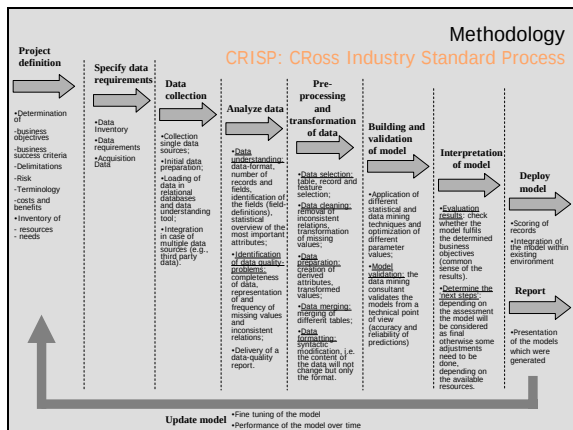
Data Mining: a non-trivial data analysis process for identifying valid, useful and understandable patterns from databases.

75.

How Data Mining Works



76.



77.

Data Mining Tasks

- Four main operations include:
 - Predictive modeling
 - Database segmentation
 - Association discovery
 - Characterization
- There are recognized associations between the applications and the corresponding tasks.
 - e.g. Direct marketing strategies use database segmentation.

78.

Data Mining Techniques

- Techniques are specific implementations of the data mining operations.
- Each operation has its own strengths and weaknesses.
- Data mining tools sometimes offer a choice of operations to implement a technique.

79.

Predictive Modeling

- Similar to the human learning experience
 - uses observations to form a model of the important characteristics of some phenomenon.
- Uses generalizations of 'real world' and ability to fit new data into a general framework.
- Can analyze a database to determine essential characteristics (model) about the data set.

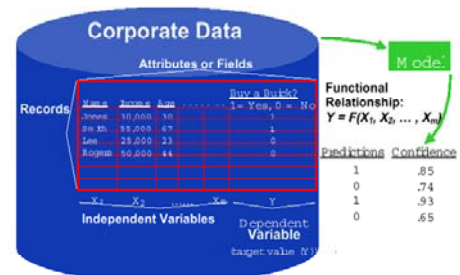
80.

Predictive Modeling

- Model is developed using a supervised learning approach, which has two phases: training and testing.
 - Training builds a model using a large sample of historical data called a training set.
 - Testing involves trying out the model on new, previously unseen data to determine its accuracy and physical performance characteristics.

81.

Predictive modeling vs scoring



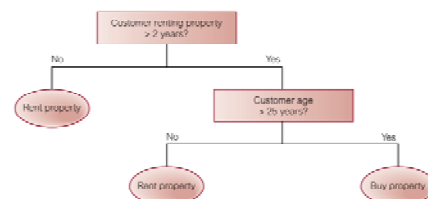
82.

Predictive Modeling

- Applications of predictive modeling include customer retention management, credit approval, cross selling, and direct marketing.
- Two techniques associated with predictive modeling: classification and value prediction, distinguished by nature of the variable being predicted.

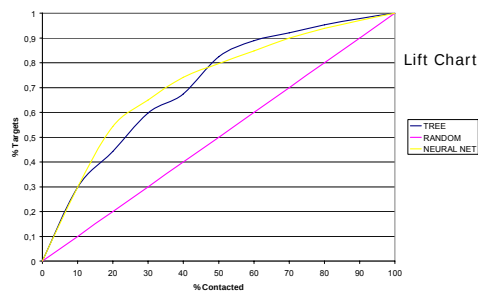
83.

Example of Classification using Tree Induction



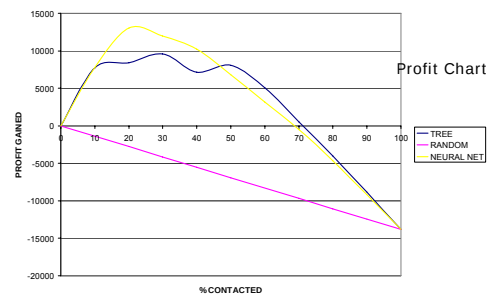
84.

Pred.modeling : Lift Chart



85.

Predictive modeling : Profit Chart



86.

Predictive Modeling - Value Prediction

- Data mining requires statistical methods that can accommodate non-linearity, outliers, and non-numeric data.
- Applications of value prediction include credit card fraud detection or target mailing list identification.

87.

Database Segmentation

- Aim is to partition a database into an unknown number of segments, or clusters, of similar records.
- Find homogeneous groups of clients who possess a similar purchasing behaviour



88.

Segmentation proces

- Identify segments (clusters)
- Select target segments
- Profile target segments
- Develop services/products to target segments
- Put an attractive (for the target segments) service/product on the market

89.

Segmentation

- Segments should be identifiable, substantial, within reach, stable in time, response sensitive
- Segments are based on geographical, socio-demo, product portfolio, product usage, payment, ... information

90.

Segmentation

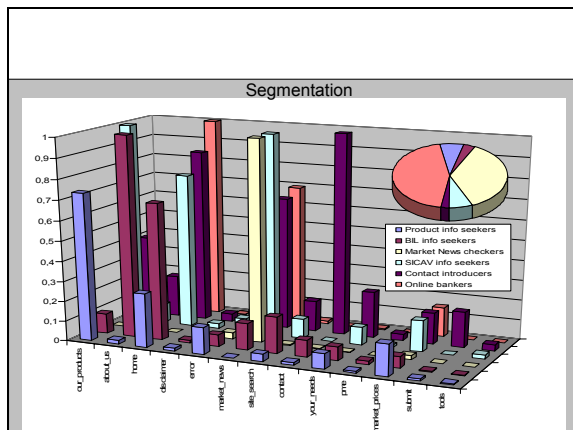
- A priori segmentation
 - Segments are given
 - Analyse and profile segments on different basis
- Posthoc segmentation
 - Determine target/profile variable
 - Cluster
 - Analyse and profile clusters

91.

Usage segmentation

Cluster	Financial	Bank	TV	Radio	News
0-0	30	143	25	5	93
0-1	1	5	25	18	9
0-2	1	3	223	23	8
1-0	13	0	1	0	0
1-1	23	0	0	0	0
1-2	9	2	6	1	3
2-0	164	1	9	2	17
2-1	38	3	1	0	19
2-2	170	14	9	8	100

92.



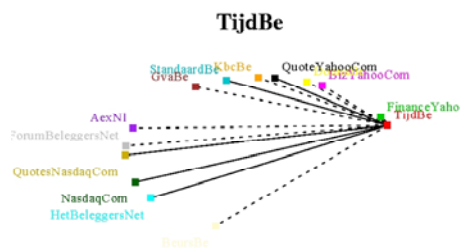
93.

Associations Discovery

- Finds items that imply the presence of other items in the same event. By discovering popular product combinations, latent interests in product associations can be found and used to target customers with the right product offers (target-selling and up-selling).
- Affinities between items are represented by association rules.
 - e.g. 'When customer rents property for more than 2 years and is more than 25 years old, in 40% of cases, customer will buy a property.'

94.

Example: www.tijd.be



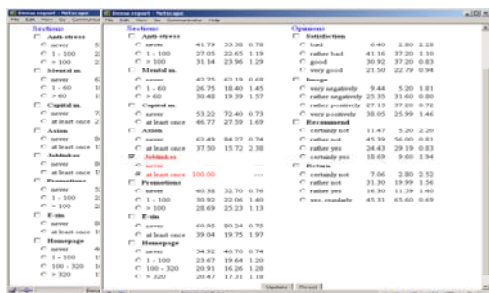
95.

Sequential Pattern Discovery

- Finds patterns between events such that the presence of one set of items is followed by another set of items in a database of events over a period of time.
 - e.g. Used to understand long term customer buying behavior.

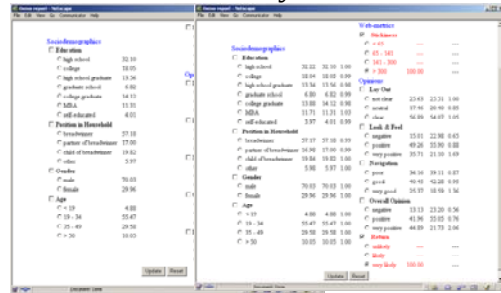
96.

Characterization



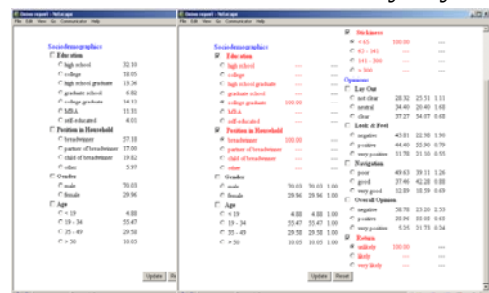
97.

Char. : Most loyal audience



98.

Possible reasons for disloyalty



99.

Outline

Explain the reason for different business intelligence technologies, show the added value and give best practices

- Datawarehousing
- Olap – multidimensional databases
- Data mining
- Integration

100.

Data Mining and Data Warehousing

- A data warehouse is well equipped for providing data for mining.
- Data quality and consistency is a pre-requisite for mining to ensure the accuracy of the predictive models. Data warehouses are populated with clean, consistent data.

101.

Data mining and Olap

- Segmentation results are the elements of a dimension
- Prediction/classification elements defines the structure of the Olap cube (which dimensions) and add metrics
- Characterization/dependency analysis defines dimensions and the appropriate metrics

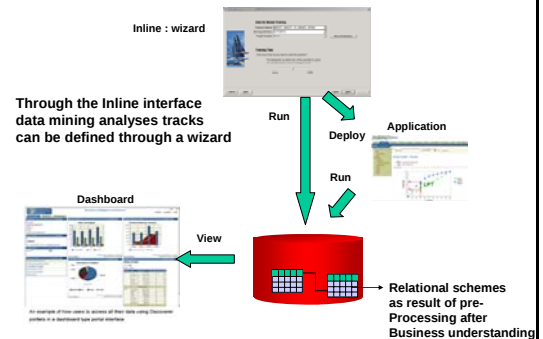
102.

Functional integration Standards

- Integration:
SQL2000
Oracle 9i
IBM DB2
- Standards : PMML, Java API, ...

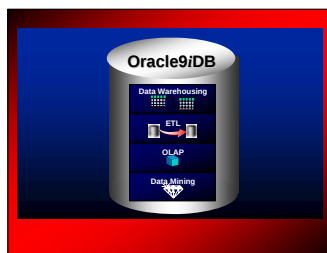
103.

Functional Integration



104.

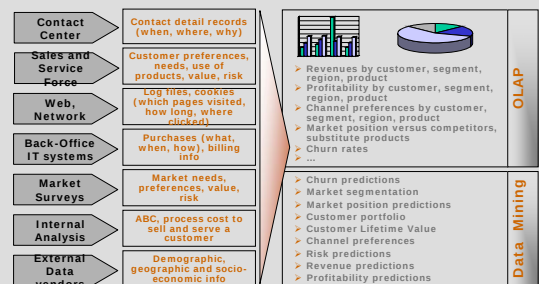
Oracle Business Intelligence Vision



- Single database
- Single server
- Single engine
- Standard interfaces
- Simplified environment
- Fastest conversion of data to information

105.

Conclusion BI : DW, OLAP, DM



106.

Author's Information

Prof. Dr. **Koen Vanhoof** - Limburgs Universitair Centrum; Belgium, koen.vanhoof@uhasselt.be

THE ENTREPRENEURSHIP IN INFORMATION AND TELECOMMUNICATION TECHNOLOGIES – AN EXPERIENCE FROM ONE COURSE IN BERKELEY

Neli Maneva

Abstract.

The main goal of the lecture is to reveal the essence of the entrepreneurship and its role for economic development and prosperity in each country. At the beginning some alarming symptoms of a crisis in the area of the ICT have been analysed. One possible solution can be the entrepreneurship – the pursuit of opportunity beyond the resources you currently control. A number of statements, supporting this approach from different points of view, is presented. After that a brief description of the entrepreneurial process is given.

Next follows the presentation of the educational programme developed by Intel and the University of Berkeley. The aim of their joint project is to support Universities, Education institutions and Governments in their endeavor to create entrepreneurial eco-systems, promoting innovation and creating the basis for economic development. The objectives of the three programs, developed till now, have been discussed. The proposed layers, structure and content of the flagship entrepreneurship course have been presented in details.

At the end of the lecture some lessons learned during the Global Faculty Colloquium are shared in two different directions: methodological and organizational. Finally, the impact of the experience gained during the Colloquium on some future educational and research activities of the author will be mentioned.

Author's Information

*Prof. Dr. **Neli Maneva** – Software Engineering Dept., Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str. bl.8, 1113 Sofia, Bulgaria; e-mail: neman@gbg.bg*

ЭЛЕКТРОННЫЙ РЫНОК ЗНАНИЙ

Красимир Марков, Красимира Иванова, Илия Митов

Аннотация: В лекции рассматриваются предпосылки возникновения и развития электронного рынка знаний, который постепенно занимает свое место среди всех форм обмена знаниями.

Введение

Одной из важнейших предпосылок существования и развития человеческого общества является обучение. Знания, умения и навыки не передаются человеку на генном уровне, т.е. по наследству. Поэтому существование и развитие человечества непосредственно связано с процессами получения, хранения, передачи и использования знаний. Известно, что есть зависимость между уровнями грамотности отдельных наций и количеством валового внутреннего продукта на душу населения тех самых наций, которые являются самой достоверной приметой развития данного общества. Как правило, после определенного периода времени ведущие нации в области образования становятся ведущими и по количеству валового внутреннего продукта на душу населения.

На современном этапе развития происходит процесс вложения нового смысла в термин “грамотность”, он становится все более практически направленным. В начале прошлого века в это понятие вкладывалось то, что люди могли “читать” и “писать”. В семидесятых годах, акцент был перенаправлен на “понимание”. Не прошло и двадцати лет, как уже данный человек считался грамотным, т.е. мог чувствовать себя адекватно и комфортно в новой среде, если умел “применять” накопленные знания в практической деятельности. Теперь это недостаточно. Основной приметой развития современного общества является **коммуникация**. Весь наш мир охвачен этим явлением. И поэтому “способность коммуницировать” является основным критерием оценки грамотности в наши дни [Vamosh, 2001].

Каждый человек стремится к знаниям, чтобы использовать их для улучшения стандарта своей жизни, что прямо связано с желанием найти подходящую работу и продвижение в ней. Таким образом, человек достигает оптимального использования своих возможностей с целью достижения достойного общественного положения [Barneva, Barnev, 2001].

Современное общество находится в корне на другом уровне развития по сравнению с прежними этапами своего функционирования. Впервые за тысячелетнюю историю человечества существование некоторых профессий становится короче, чем продолжительность трудовой деятельности отдельного человека. Это явление трудно осознается, а люди пока еще не подготовлены к нему.

Схема “впервые очередь получить образование, а потом использовать накопленные знания в профессиональной работе” становится *неуместной*. В настоящее время более правильным является осуществление построения образовательного процесса по другой схеме:

- первоначальное усвоение некоторого объема фундаментальных знаний и умений, которые являются валидными для продолжительного периода времени;
- систематическое получение актуальных знаний и умений (на базе интенсивных курсов), которые будут сразу применяться, но недолго.

Существует принципиально ошибочное и непродуктивное отношение к образованию, которое часто выражают фразой: “дали бы нам образование”. Образование требует высокой активности и сильных волевых качеств. Воспитание этих качеств должно начать еще в раннем (дошкольном) возрасте. Родители и учащиеся, которые думают, что школа и государство сделают их дело, идут по совершенно ошибочной и безрезультатной дороге [Barneva, Barnev, 2001].

Современное образование классифицируют, как базовое и дополнительное [Ivanova et al, 2001]:

- **Базовое образование** предполагает культурное, социальное и профессиональное развитие человека, необходимое ему и обществу. Оно включает:
 - **общее образование**, которое является фундаментальной основой, необходимой человеку для дальнейшего выбора и получения конкретного профессионального образования;
 - **профессиональное образование**, которое характеризуется определенными требованиями к знаниям, умениям и навыкам, необходимыми специалисту в конкретной сфере труда.
- **Дополнительное образование** характеризуется расширением или повышением уровня знаний, умений и навыков базового или профессионального образования для дальнейшего культурного, социального и профессионального развития человека.

В настоящее время в мире наблюдаются тенденции развития открытых образовательных структур. В Англии, США, Канаде, Швеции, Финляндии, Испании и других странах давно созданы и уже много лет развиваются **открытые университеты**, которые предоставляют возможность любому человеку получить широкий спектр образовательных услуг, в том числе, на основе дистанционных методов обучения. Стоимость такого обучения составляет в среднем 200 долларов США за 12-15 недель обучения, что является доступным для большинства жителей из экономически развитых стран ЕС или США, Японии, Канады и т.д. [Иванова и др., 2001].

Современный этап развития человечества характеризуется широким внедрением и использованием сетевых информационных технологий во всех сферах деятельности общества. Этот этап можно назвать началом глобальной информатизации человечества. Соответственно, система образования должна отражать, поддерживать и обеспечивать современные требования развития общества [Марков и др., 2000].

Требования массового и непрерывного повышения уровня профессиональной квалификации специалистов является важной особенностью развития информационного общества. Опыт развитых стран, широко использующих информационные технологии, показывает, что массовое обучение и переподготовка специалистов требует существенного использования бюджетных средств государств и отдельных регионов на социальные программы. Например, Финляндия ежегодно в конце 20-го и начале 21-ого века использует 12-16% бюджетных средств на образование и 10-14% на социальные программы, предусматривающие переподготовку специалистов. При этом крупные фирмы и предприятия имеют учебные центры, которые решают задачи подготовки и переподготовки кадров в соответствии с требованиями развития информационных технологий на этих предприятиях и в соответствии с собственными финансовыми возможностями. Опыт Финляндии показывает, что традиционные формы обучения не всегда могут полностью удовлетворить потребности общества в массовой и непрерывной переподготовке специалистов. Эффективная реализация этих потребностей связана с созданием, развитием и использованием новых педагогических подходов, новых технологий обучения и новых организационных принципов построения системы профессионального продолженного обучения взрослых [Иванова и др., 2003], учитывающие такие факторы как:

- массовый и непрерывный характер профессионального продолженного обучения;
- доступность и эффективность обучения;
- численная ограниченность преподавателей;
- большое многообразие и частое изменение учебных курсов;
- территориальная удаленность многих учеников от учебных центров и университетов.

Очевидно, традиционные образовательные формы меняются и новые элементы включаются в общий процесс. Одним таким элементом является информационный рынок, важной частью которого составляет рынок знаний [Марков и др. 2005]. Об этом поговорим в этой лекции. Лекция основана на ряда публикаций авторов, некоторые из которых указаны в литературе.

Открытый рынок дистанционных услуг

Развитие глобальных телекоммуникационных линий связи и сетевых информационных технологий подготовили условия для создания и организации дистанционно-связанных рабочих мест, использующих общую сеть компьютерного и телекоммуникационного оборудования. Это способствует созданию и развитию дистанционно объединенных предприятий, решающих общие задачи производства, исследований, торговли, медицинского обслуживания и так далее, вне зависимости от места их расположения. Происходит процесс глобального объединения рынка производства и рынка труда на основе сетевых информационных технологий. Соответственно увеличиваются конкуренция и требования к уровню квалификации специалистов. Конкуренция на глобальном сетевом рынке труда является мощным стимулом для непрерывного повышения квалификации специалистами из разных регионов и стран. Следовательно, современная система профессионального продолженного обучения взрослых должна быть так же сетевой, чтобы быстро и эффективно удовлетворять запросы в повышении квалификации специалистов из удаленных регионов. Это означает, что с развитием сетевых информационных технологий, процессов глобального объединения рынков производства и труда, в обществе появились реальные предпосылки для возникновения открытого сетевого рынка профессиональных образовательных услуг в сфере продолженного обучения взрослых [Danilov, Ivanova, 1999].

Процессы объединения на основе сетевых информационных технологий происходят и в сфере профессионального образования. Например, в Финляндии создается и развивается единая информационная сеть университетов, в ЕС создана ассоциация университетов и учебных центров, объединенных информационной сетью для решения общих задач обучения. Многие университеты и учебные центры различных государств начинают оказывать услуги для профессионального продолженного обучения взрослых на основе сетевых дистанционных методов обучения.

Таким образом, сетевые информационные технологии начинают реально находить применение в сфере профессионального образования и формируют **открытый рынок** дистанционных услуг профессионального обучения. Региональная, национальная и международная интеграция университетов и учебных центров на основе открытых сетевых информационных технологий способствует развитию общего информационного пространства в сфере образования.

Опыт развития открытого сетевого дистанционного обучения в университетах и учебных центрах разных стран показывает, что такие технологии могут обеспечить:

- массовый и непрерывный характер профессионального продолженного обучения;
- удаленный доступ к учебным материалам, преподавателям и студентам;
- интерактивность процесса обучения;
- эффективное управление учебным процессом на основе оптимального сочетания индивидуальных и коллективных методов дистанционного обучения.

Однако, необходимо отметить, что сетевые технологии дистанционного обучения имеют существенные отличия от традиционных и требуют:

- использования сетевой учебной среды, в которой происходит дистанционный интерактивный учебный процесс и экспертная оценка его результатов;
- подключения компьютеров дистанционных студентов и преподавателей к сетевой учебной среде;
- специализированных программных средств и инструментов, используемых при подготовке сетевых учебных курсов и учебных материалов;
- разработки дистанционных учебных курсов, учебных материалов, тренажеров и учебных исследовательских работ, адаптированных к сетевой учебной среде;

-
- обучения преподавателей и тьюторов новым педагогическим подходам, обеспечивающим эффективные прямые и обратные связи со студентами в сетевой учебной среде для проведения дистанционного интерактивного обучения;
 - эффективных методов дистанционного контроля и оценки знаний студентов;
 - специалистов для технической поддержки и обеспечения компьютерной безопасности сетей и баз данных открытых сетевых университетов и учебных центров;
 - Web дизайнеров для технической поддержки преподавателей при подготовке сетевых учебных материалов;
 - персонала менеджеров и координаторов для обеспечения управленческих функций сетевым университетом и информационных связей с удаленными клиентами.

Источники финансирования

Указанные отличия показывают, что сетевые технологии обучения предполагают существенные финансовые затраты при создании открытых сетевых университетов и учебных центров. Поэтому важными моментами для развития современных технологий обучения являются не только широкие возможности удаленного доступа к учебному заведению и дистанционного интерактивного процесса обучения, но и экономическая эффективность таких технологий обучения. К основным источникам возможного финансирования открытых сетевых университетов и учебных центров можно отнести:

- ресурсы бюджета (местный, региональный и государственный);
- финансовые ресурсы предприятий и фирм (часть стоимости продукции или часть прибыли);
- различные фонды и программы (государственные, региональные, частные, общественные, национальные или международные);
- финансовые ресурсы студентов (личные или кредит).

Конечно, обычно пользуются смешанные финансовые ресурсы.

Рынок знаний

В основе экономической эффективности открытых сетевых университетов и учебных центров должен быть **принцип самокупаемости** вне зависимости от источников финансирования. В противном случае сетевые дистанционные технологии обучения не будут иметь широкого развития из-за нехватки финансовых ресурсов. Поэтому одной из основных задач команды управления открытым сетевым университетом или сетевым региональным учебным центром является изучение и оценка рынка потенциальных потребителей дистанционных учебных курсов. Подготовка конкретного учебного курса должна начинаться только тогда, когда определен рынок потенциальных потребителей и сделана оценка экономической эффективности обучения. Необходимо заметить, что потенциальными потребителями сетевых дистанционных курсов могут быть студенты из разных регионов или государств. Следовательно, современные технологии сетевого дистанционного обучения являются технологической основой для формирования глобального **открытого рынка** дистанционных образовательных услуг. Эти технологии дают возможность студентам, на платной основе, учиться в Открытом сетевом дистанционном университете любой страны, не пересекая границу своего региона и государства.

Финансовые взаимоотношения основных участников **рынков знаний** базируются на договорных обязательствах между покупателем и продавцом знаний, которые определяют услуги обучения и их стоимость в соответствии с юридическими правилами и законами, защищающими права покупателя и продавца знаний. Рынки знаний представляют собой организованную совокупность участников учебных сред, действующих по единым правилам и законам, регулирующим эти рынки, причем каждый участник в зависимости от потребности рынка может быть продавцом или покупателем знаний.

Структура рынка знаний

Основная задача при анализе среды знаний – выявление ее основных компонент и взаимосвязи между ними – структура среды знаний формируется из совокупности взаимно-связанных элементов, функционирующих одновременно во времени и разделяющие общие ресурсы. На базе анализа существующих подходов сбора, сохранения, переработки и распространения знания и учитывая основные характеристики открытой среды знания мы можем построить обобщенную схему открытой среды знаний, отражающая информационные связи и взаимодействия участников среды знания [Markov et al, 2002].

Первый вопрос, на который необходимо ответить при анализе - "Кому необходимы знания, и кто является реальным покупателем знаний?" Вероятнее всего, что таковым является тот, кто реально на практике занимается предпринимательством, т.е. имеет и развивает определенный бизнес. Для развития и дальнейшего процветания этого бизнеса он нанимает люди, у которых есть определенные знания и умения и в процессе производства превращают свои знания в материальные объекты, за которое работодатель платит соответствующее вознаграждение. В этом взаимодействии предприниматель является в роли **Работодателя (Р)**, а исполнители деятельности являются в роли **Служащих (С)**. Оно обслуживается рынком труда.

То есть, весь процесс начинается от Работодателя. Поэтому Работодатель есть первая компонента, а второй компонентой в системе являются Служащие - т.е. люди, которые владеют знаниями и могут своим трудом превратить их в реальные объекты.

Если бы люди рождались генетически всезнающие, тогда бы систему можно было бы замкнуть до этих двух компонент. Однако люди, даже имеющие высокий уровень образования, нуждаются в дополнительных знаниях, для того, что бы выполнять заказы Работодателя. Именно в этот момент они превращаются в Покупателями новых знаний. Это не знания, полученные от преподавателей в средней или высшей школах, а дополнительные знания, которые необходимы для выполнения новых заказов Работодателя.

Необходимость приобретения новых знаний порождает новый рынок – рынок знаний, который должен быстро реагировать и обслуживать нужды покупателей знаний. Т.е. рынок труда причиняет появления рынка знаний, конечный результат цикла которого – обученных и опытных работников – покупает Работодатель.

Технологический и социальный статус общества не статичный. Он постоянно изменяется – открываются новые знания, на базе которых развиваются новые технологии. Чтобы эти технологии достигли до производства, нужно довести их до знания работодателя – заинтриговать его до степени внедрения в его деятельности (производства) с одной стороны и определить способов обучения кадров для обслуживания новых технологией с другой стороны. Так оформляется специальный клас участников на рынке знаний – **Консультанты (К)**. Это люди или институты (научные, бизнес, общественные или государственные), которые специализируются в определенной практической или научной области и проводят исследования по применению новых знаний и технологий в различных сферах деятельности.

Сами Консультанты не должны заниматься обучением кадров. Они только определяют что кадры должны знать. Обучение проводится **Преподавателями (П)**, которые трансформируют научные знания, определенные Консультантом в педагогически обоснованные лекции и занятия.

Преподаватель имеет ограниченные ресурсы по координации процесса передачи знаний и взаимодействию с каждым покупателем знаний, по рекламе предоставляемых им услуг и т.д. Для решения этих задач, Преподаватель нуждается в помощи ассистента, который организует учебный процесс и консультирует обучаемых. Это **Тьютор (Т)**.

Работодатель не в состоянии проводить самостоятельный контроль знаний и умений своих настоящих или будущих служителей и определить степени ихней адекватности должностям, которым занимают или

будут занимать. Теоретическая обосновка адекватности задается консультантом, но реальная проверка и аттестация проводится другим участником рынка знаний – **Экзаменатор (Э)**.

Таким образом, мы получаем шесть компонент, которые должны свободно функционировать в глобальной информационной сети и быть связаны друг с другом, взаимодействуя между собой. Они формируют превыш уровень рынка знаний, которого можно назвать **"уровень информационного взаимодействия"**.

Поскольку вышеуказанные компоненты много и разбросаны в пространстве планеты для организации ихнего свободного взаимодействия необходимо соответствующее **"информационное обслуживание"**. Оно делается новой компонентой, которая координирует и связывает их между собой. Роль такой компоненты выполняет провайдер информационных услуг – **Администратор (А)**. Администратор характеризуется знанием и возможностями для осуществления и менеджирования процессов обмена информации между членами данного рынка знаний.

Нарастающая активность рынка знаний обуславливает необходимость разработки современных средств информационного обслуживания.

Для этого создаются коллективы по **"информационному моделированию"** специфических характеристик и информационных процессов рынка знаний, его целостное настоящее и будущее развитие. На базе информационных моделей разрабатываются адекватные средства информационного обслуживания. Этими деятельностью занимаются **Исследователи (Ис)** и **Разработчики (Рз)**.

Этот уровень рынка знаний позволяет посмотреть на весь этот процесс со стороны, чтобы исследовать его, изучать и развивать.

Он влияет на процесс развития системы, являясь ее внешним возмущителем.

Исходя из этих рассуждений, и прослеживая их взаимосвязей можно прийти к схеме рынка знаний (рис.1).

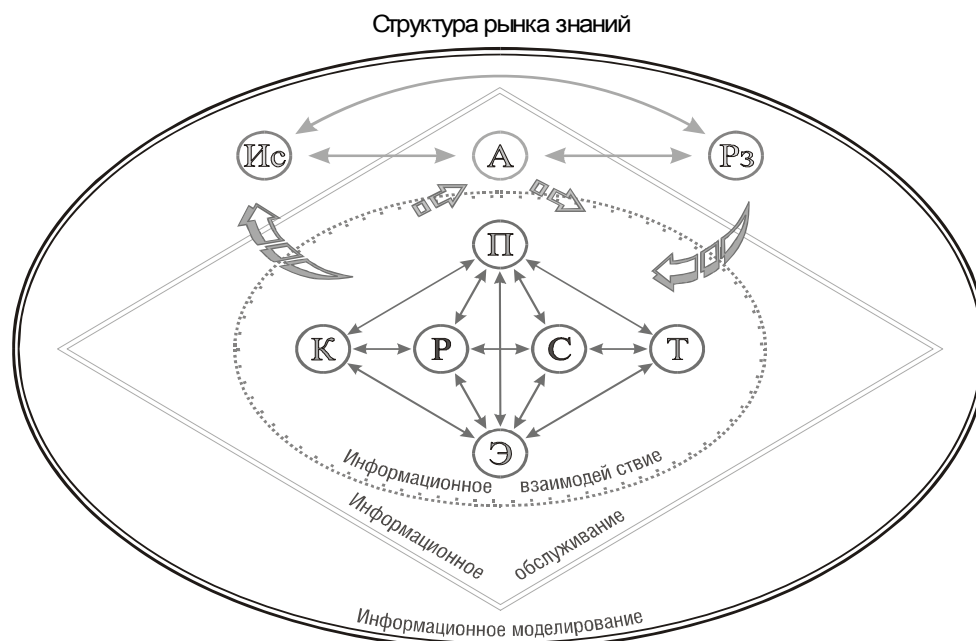


рис. 1. Структура рынка знаний

Электронный рынок знаний

Одной очень важной частью рынка знаний является электронный рынок знаний. Основа этого рынка – среда обучения, построенная на базе сетевых информационных технологий. На современном этапе развития общества электронный рынок знаний начинает доминировать. Из-за того что электронный рынок

по своей сути глобальный, в настоящем моменте практически невозможно юридически регулировать взаимоотношения между участниками электронного рынка знаний средствами отдельного государства. В результате необходимо достичь договоренности на международном уровне по регулированию взаимоотношений на открытом электронном рынке знаний. Существующие в настоящем моменте международные правовые нормы являются хорошей начальной юридической структурой, усовершенствование которой крайне необходимо и важная задача близкого будущего.

Подготовка услуги сетевого дистанционного обучения непосредственно связана с приобретением или созданием сетевой среды, в которой будет происходить процесс обучения. Такую информационно-технологическую среду обучения можно назвать виртуальный класс, виртуальный колледж, виртуальный университет и т.п. В настоящее время разработаны и используются много различных сетевых сред обучения.

Необходимо отметить, что любые сетевые учебные среды должны удовлетворять следующим основным требованиям учебного процесса:

- открытость только для студентов и преподавателей курса обучения;
- защита баз данных преподавателей от несанкционированного доступа и возможного разрушения;
- защита баз данных студентов от несанкционированного доступа и возможного разрушения.

Это означает, что учебная среда, даже в открытой глобальной системе обучения не является полностью открытой. Ограничения необходимы для обеспечения защиты учебного процесса от внешних и внутренних негативных действий, которые могут нарушить основные принципы дистанционного обучения. Заметим, что эти ограничения при использовании сетевых дистанционных технологиях обучения имеют свои аналоги при традиционном обучении в классических учебных заведениях. Следовательно, эти ограничения являются принципиально необходимыми для построения эффективного процесса обучения. Сетевая учебная среда должна помогать преподавателям и студентам построить наиболее оптимальную стратегию сетевого дистанционного обучения. Эту функцию можно реализовать на основе использования искусственного интеллекта для управления учебным процессом. Такие требования к функциональным возможностям сетевой учебной среды становятся особенно важными при массовом обучении студентов, когда имеются реальные ограничения на количество преподавателей и тьюторов. В этом случае интеллектуальная система управления обучением является необходимым помощником преподавателей и студентов.

Учебные материалы и тренажеры, адаптированные к сетевой учебной среде являются необходимыми и наиболее дорогостоящими составляющими дистанционного учебного процесса [Марков и др., 1998]. Это объясняется тем, что в одной и той же сетевой учебной среде можно проводить обучение по самым различным курсам, но каждый учебный курс должен быть обеспечен своими учебными материалами и приложениями. Использование при их подготовке мультимедийных возможностей информационных технологий, технологий виртуальной реальности и возможностей искусственного интеллекта для достижения наибольшей эффективности дистанционного обучения делают такие учебные материалы дорогостоящими. Соответственно только массовое использование этих учебных материалов может оправдать финансовые затраты на их разработку и реализацию. Поэтому при подготовке учебных материалов для нового сетевого дистанционного курса обучения необходимо использовать наиболее опытных преподавателей и педагогические подходы, которые показали свою эффективность [Boikatchev et al, 2001a]. Надо отметить, что педагогические подходы при использовании сетевых дистанционных технологий обучения в отличие от классических методов обучения должны дополнительно учитывать:

- технические и технологические возможности сетевых информационных технологий для дистанционного интерактивного общения со студентами и подготовки учебных материалов;
- национальный менталитет студентов в многонациональной дистанционной учебной группе;

- необходимость поддержки удаленных прямых и обратных связей с каждым дистанционным студентом в течении всего периода обучения.

Поэтому одной из важных задач системы образования в информационном обществе является подготовка преподавателей и тьюторов для работы в новых условиях рынка знаний [Boikatchev et al, 2001b].

Защита авторских прав

Очень важным фактором для развития открытого рынка знаний является защита интеллектуальной собственности и авторских прав на сетевые дистанционные учебные материалы. Известно, что не существует абсолютно надежной защиты от копирования интеллектуальных продуктов. Поэтому использование сетевых дистанционных учебных материалов без разрешения авторов должно подвергаться соответствующему наказанию. Для реализации этих требований должна быть создана международная сетевая служба защиты от незаконного использования авторских сетевых дистанционных учебных материалов. Кроме этого должна быть международная сетевая служба, которая разыскивала и наказывала хакеров, которые принесли ущерб открытым сетевым университетам, центрам и их студентам, в результате несанкционированного доступа и разрушения учебной сети. Это означает, что открытость сетевого дистанционного обучения требует общей международной защиты [Марков, 1999].

Специализированные поисковые системы

Развитие открытого сетевого рынка дистанционных услуг обучения невозможно без информационной поддержки в традиционных средствах массовой информации и Internet. Распространение информации о возможностях сетевого дистанционного обучения в различных университетах и учебных центрах через Internet является наиболее перспективным в информационном обществе. Практика показывает, что современные университеты и учебные центры имеют веб сайты, на которых расположена информация о предоставляемых услугах обучения. Однако это не означает, что эту информацию можно легко найти в Internet, и она будет полностью достоверной. Поэтому для эффективного решения этой задачи необходимо создать специализированные интеллектуальные поисковые системы, которые должны обладать следующими возможностями:

- ввод запроса о поиске информации об образовательных услугах на любом языке;
- осуществлять поиск информации во всех языковых ресурсах Internet и предоставлять ее пользователю на языке запроса;
- иметь возможность организовать связь с менеджерами образовательных услуг университетов или учебных центров и в интерактивном диалоге уточнить достоверность найденной информации;
- иметь интерактивный интерфейс с пользователем для ранжирования результатов поиска информации по критериям пользователя;
- интеллектуально обрабатывать найденную информацию об образовательных услугах и делать экспертное заключение о результатах поиска.

Самостоятельная работа обучающегося

Важной особенностью современных дистанционных технологий обучения является большая самостоятельная работа студента с учебными материалами. Соответственно у студента должны быть навыки самостоятельной работы и мотивация к получению новых знаний при большом объеме самостоятельного изучения различных учебных материалов. Этим требованиям в наибольшей степени удовлетворяют взрослые люди, которые хотят получить дополнительное образование для повышения профессионального уровня и успешной реализации профессиональной карьеры. Кроме этого, взрослые люди в информационном обществе представляют огромный рынок потенциальных потребителей образовательных услуг продолженного обучения. Таким образом, продолженное или дополнительное

профессиональное обучение взрослых в информационном обществе является мощным стимулом для развития открытых сетевых дистанционных технологий обучения и электронного рынка знаний.

Необходимо заметить, что продолженное обучение взрослых в информационном обществе может эффективно осуществляться на основе самостоятельного обучения специалиста без получения образовательных услуг от университетов или региональных учебных центров. Этому способствуют опыт самостоятельной профессиональной работы специалиста, мотивация взрослого человека к получению профессиональных знаний и возможность удаленного доступа к открытым сетевым информационным ресурсам. Увеличение общедоступной и разнообразной информации является основным признаком развития открытого информационного общества. Открытые сетевые информационные возможности традиционных и виртуальных библиотек являются основными учебными материалами для самостоятельного обучения взрослых, которые можно найти в Internet. Компьютерные или электронные копии традиционных книг и статей, а так же информационные ресурсы виртуальных библиотек позволяют использовать возможности искусственного интеллекта для поиска, обработки и оценки желаемой информации в этих информационных источниках.

В то же самое время, достижение необходимого уровня знаний для дальнейшего самообучения нелегкая задача, в решении которой нужно подключить ресурсы регионального центра электронного рынка знаний, включительно индивидуальное дистанционное или директное обучение.

Заключение

Электронный рынок знаний постепенно занимает свое место среди всех форм обмена знаний. Платная форма вероятно будет ведущей в нынешних условиях. Выше мы обсудили только некоторые предпосылки возникновения и развития электронного рынка знаний. Предстоит еще много работы по исследованию этого феномена информационного общества.

Подчеркнем, что в информационном обществе появляется возможность интеграции традиционных и сетевых дистанционных технологий обучения на основе использования удаленного сетевого доступа к библиотечным ресурсам, сетевым учебным материалам и их интеллектуальной обработке для более эффективного использования. В развитом информационном обществе традиционные методы обучения будут использовать все возможности сетевых дистанционных технологий для повышения эффективности образования. Например, сетевые тренажеры или сетевые учебные исследовательские работы будут использоваться как дистанционные технологии в традиционном обучении. Стоимость создания таких учебных материалов может быть весьма высокой. Однако огромное количество пользователей из разных регионов страны или из разных стран позволят сделать такие разработки экономически эффективными. Поэтому учебные материалы будут создаваться на основе самых высоких сетевых информационных технологий, и ими будут пользоваться как традиционные, так и открытые сетевые колледжи и университеты. Трехмерная сетевая интерактивная видеосвязь предоставит возможность удаленным студентам получать учебные услуги, которые по качеству не будут отличаться от учебных услуг, получаемых при традиционном обучении. Эта техническая возможность окажет существенное влияние на формирование учебных курсов и преподавательского состава, как традиционных, так и открытых сетевых университетов. Конкуренция на рынке образовательных услуг традиционных университетов в открытом информационном обществе потребует от них улучшения качества и расширения возможностей обучения. Традиционные университеты будут развивать и использовать наиболее эффективные сетевые технологии обучения, предоставляя возможность своим студентам дополнительно учиться у лучших преподавателей из разных стран. В результате широкого использования сетевых дистанционных технологий обучения, традиционные университеты будут представлены на глобальном сетевом рынке образовательных услуг, как и открытые сетевые университеты. Это позволит создать максимальные возможности предоставления образовательных услуг на глобальном сетевом рынке обучения.

Проведенный анализ влияния начального этапа информатизации общества на возможность развития открытой системы обучения показывает:

-
- открытые сетевые информационные технологии являются техническим базисом для развития общего сетевого рынка труда;
 - конкуренция на общем сетевом рынке труда и частое обновление информационных технологий в сфере труда требуют от специалистов постоянного повышения уровня профессиональных знаний;
 - традиционные методы обучения не могут полностью обеспечить спрос сетевого рынка труда на массовые и непрерывные образовательные услуги продолженного профессионального обучения взрослых в быстро развивающемся информационном обществе;
 - сетевые технологии дистанционного обучения предоставляют возможность для организации массового продолженного профессионального обучения взрослых, которые удалены от традиционных университетов или учебных центров;
 - открытые сетевые университеты являются основой для создания открытого электронного рынка знаний;
 - открытые сетевые технологии удаленного доступа к информации и клиентам, мультимедийные информационные технологии, технологии виртуальной реальности, технологии приобретения, обработки и использования знаний, технологии искусственного интеллекта являются информационной технологической основой сетевого дистанционного обучения;
 - эффективное развитие открытого сетевого дистанционного обучения непосредственно связано с принципами самоокупаемости открытых сетевых университетов и региональных центров электронного рынка знаний;
 - педагогические подходы открытого сетевого дистанционного обучения должны учитывать возможности сетевых информационных технологий для создания учебных материалов и организации интерактивного учебного процесса, национальный менталитет дистанционных студентов из многонациональной учебной группы, необходимость поддержки прямых и обратных связей с каждым дистанционным студентом в течении всего периода обучения;
 - технологические и информационные возможности электронного рынка знаний должны помогать организовать эффективное управление массовым сетевым дистанционным обучением при ограниченной численности преподавателей;
 - открытость сетевого рынка услуг дистанционного обучения не означает полную открытость сетевой учебной среды;
 - для организации эффективного дистанционного обучения сетевая учебная среда должна обеспечивать открытый доступ только студентам и преподавателям конкретного учебного курса;
 - развитие электронного рынка образовательных услуг сетевого дистанционного обучения не возможно без создания интеллектуальных поисковых систем, которые могут обеспечить потенциальным клиентам сетевого дистанционного обучения поиск достоверной информации об услугах обучения в открытых сетевых университетах и организовать интерактивную связь с менеджерами этих университетов;
 - защита авторских прав разработчиков сетевых учебных материалов и защита от несанкционированного доступа в учебные сети открытых университетов являются важной международной задачей по обеспечению безопасности открытого электронного рынка знаний;
 - конкуренция на рынке образовательных услуг традиционных университетов способствует развитию и использованию в этих университетах наиболее эффективных сетевых дистанционных технологий обучения, которые могут предоставить традиционным студентам дополнительные возможности для получения образовательных услуг у лучших преподавателей университетов различных стран;
 - в развитом информационном обществе открытый рынок знаний будет интегрировать дистанционные возможности обучения открытых и традиционных университетов.

Литература

- [Barneva, Barnev, 2001] M. Barneva, P. Barnev. *The Education in Informatics and the Information Society*, Proceedings of 26th ICT&P, 2001, Sofia, Bulgaria, ISBN 954-16-0019-0, pp. 143-154
- [Boikatchev et al, 2001a] K. Boikatchev, N. Ivanova, A. Danilov, Kr. Markov, Kr. Ivanova. *Authoring Tools for Courseware Designing*. Int.Journal on Information Theories and Applications, 2001, V.8, No.3 pp.115-121.
- [Boikatchev et al, 2001b] K. Boikatchev, N. Ivanova, A. Danilov, Kr. Markov, Kr. Ivanova. *Teacher's and Tutor's Role in Long Life Distance Learning*. Int.Journal on Information Theories and Applications, 2001, V.8, No.4 pp.171-175.
- [Danilov, Ivanova, 1999] A. Danilov, N. Ivanova. *Knowledge-based Human Resources Management – Concept and Implementation*. Materials of International Knowledge Technology Forum KnowTechForum, Potsdam, September, 1999.
- [Ivanova et al, 2001] N. Ivanova, Kr. Ivanova, Kr. Markov, A. Danilov, K. Boikatchev. *The Open Educational Environment on the Threshold of the Global Information Society*. Int.Journal on Information Theories and Applications, 2001, V.8, No.1 pp.3-12.
- [Markov et al, 2002] Kr. Markov, Kr. Ivanova, I. Mitov, N. Ivanova, A. Danilov, K. Boikatchev. *Basic Structure of the Knowledge Market*. Int.Journal on Information Theories and Applications, 2002, V.9, No.4, pp.123-134 (Presented at Int. Conf. ICT&P, 2002, Primorsko)
- [Vamosh, 2001] T. Vamosh, *Education – the key – how different?*, Proceedings of 26th ICT&P, 2001, Sofia, Bulgaria, ISBN 954-16-0019-0, p.7
- [Иванова и др., 2001] Н. Иванова, А. Данилов, П. Вишневский, Й. Иммонен. *Опыт развития современных образовательных технологий в Финляндии*. Тезисы IV межрегиональной научно-практической конференции «Проблемы и перспективы взаимодействия вузов Санкт-Петербурга с регионами России в контексте реформирования образования». Март 2001 г. Смольный.
- [Иванова и др., 2003] Кр. Иванова, Н. Иванова, А. Данилов, И. Митов, Кр. Марков. *Обучение взрослых на рынке профессиональных знаний*. Сборник доклады на Национална научна конференция “Информационни изследвания, приложения и обучение” (i.TECH-2003), Варна, България, 2003. Стр. 35-41.
- [Марков, 1999] Кр. Марков. *Относно вредите от електронната търговия*. IECC'99: International e-Commerce Conference. ADIS & VIA EXPO, Bulgaria, Sofia, 1999.
- [Марков и др., 1998] Кр. Марков, Кр. Иванова, И. Митов. *Исследования информационной поддержки образования*. НИТЭ-98: Новые информационные технологии в преподавании электротехнических дисциплин - Изд-во АГТУ, Астрахань, 1998, стр. 51-56.
- [Марков и др., 2000] Кр. Марков, Кр. Иванова, И. Митов. *Требования к автоматизации обучения на пороге информационного общества*. Новые информационные технологии в электротехническом образовании (НИТЭ-2000): Сборник научных трудов пятой международной научно-методической конференции. – Россия, Астрахань: Изд-во ЦНТЭП, 2000.
- [Марков и др., 2005] Кр. Марков, Н. Иванова, А. Данилов, Кр. Иванова, И. Митов. *Электронный рынок знаний: предпосылки возникновения и развития*. Сб. „Моделирование устойчивого регионального развития”, Россия, Нальчик, 2005, стр. 35-46.

Информация об авторах

Крассимир Марков: Институт информационных теории и применений ФОИ ИТЕА, Институт математики и информатики, БАН, Болгария; e-mail: markov@foibg.com

Крассимира Иванова: Институт математики и информатики, БАН, Болгария; e-mail: kivanova@math.bas.bg

Илия Митов: Институт информационных теории и применения ФОИ ИТЕА, Болгария; e-mail: mitov@foibg.com

60th Anniversary of



**INSTITUTE OF
MATHEMATICS AND INFORMATICS
of Bulgarian Academy of Sciences**

Acad. G. Bonchev Str., block 8, Sofia 1113, Bulgaria

Tel. (+359-2) 979-3824, Fax (+359-2) 971-3649

<http://www.math.bas.bg>

The Institute of Mathematics and Informatics (IMI) at BAS was founded in 1947 as Institute of Mathematics. At the beginning about ten research fellows were working at the Institute. In 1961 a computational centre was established as part of the Institute. Later specialist in Mechanics also worked at the Institute, hence and it was named Institute of Mathematics and Mechanics. Its present name dates from 1995.

The Institute has considerable achievements in the field of Mathematics that are not discussed here.

The development of the Informatics in Bulgaria started at the Institute. Many researchers have built the career of Informatics specialists.

The Institute was the first in Bulgaria to buy an universal analog computing machine MH-7. The first Bulgarian computer was created at the Institute. Soon after that came into exploitation the first imported into Bulgaria computer "MINSK-2". An original software for this computer – auto code "MIKOD", operation systems "MID" and "MID-2", a system for symbol programming "MIKS" and a rich library of programs were created here as well.

The fellows of the Institute also carried out the first Informatics researches in Bulgaria. The Institute has a wide range of activities in Applied Informatics and it continues to produce original software for the solving important problems. Researchers from the Institute organized and taught the first courses in Informatics at the Sofia University "St. Kliment Ohridski" for students in Mathematics. In a short time a major in Informatics was launched with the help of the Institute and later on it became a specialty at the Sofia University. Researchers of the Institute prepared the first syllabus, textbooks, and manuals. The staff of the Institute is also involved in training teachers in Informatics for the secondary school.

In the course of the years the informaticians at focused upon the research activities and many of them are still lecturing Informatics at a number of Bulgarian universities.

Departments of IMI : *Algebra; Artificial Intelligence; Biomathematics; Complex Analysis; Differential Equations; Education in Mathematics and Informatics; Geometry and Topology; Information Research; Laboratory of Mathematical Linguistic; Logic; Mathematical Foundations of Informatics; Mathematical Linguistics; Mathematical Physics; Computational Mathematics; Operation Research; Probability and Statistics; Real and Functional Analysis; Software Engineering; Telecommunications Department.*

