BELARUSIAN STATE UNIVERSITY

On the rights of the manuscript

UDC 004.8:519.816 (043)

VISSIA
Herman Elisabeth Rene Maria

MODELS, ALGORITHMS AND THE TECHNOLOGY FOR DECISION MAKING INTELLECTUALIZATION BASED ON SUBJECT COLLECTIONS

Dissertation for the scientific degree of PhD in technical sciences

Speciality 05.13.17 – “Theoretical Foundations of Computer Science”

Scientific adviser:

Doctor of technical sciences, professor Krasnoproshin V.V.

Minsk 2012
SHORT INTRODUCTION

Decision making intellectualization is currently an important issue in state policy of developed countries. Problems of intellectualization are reflected in the corresponding government documents and programs, including “Framework Programs for Research and Technological Development 2011-2012” (EC), “Government ICT Strategy” (Great Britain), “Computational Science: Ensuring America’s Competitiveness” (USA), etc.

“Intellectualization” is the involvement of innovative knowledge, which should ensure competitive advantages for government agencies, private companies and various sectors of the population. The main sources of such knowledge are experts, dispersed in leading scientific and technological centers of the world. The primary means of intellectualization are computer Decision Support Systems (DSSs), which accumulate and distribute the innovations.

Theoretical and technological problems of decision making intellectualization are investigated in works of well-known scientists: R.Ackoff, R.Brachman, L.Byars, C.Christensen, A.Brooking, I.Dahlberg, D.DeCenzo, A.Doan, P.Drucker, J.Galbraith, D.Luger, I.Palmer, Yu.Zhuravlev, L.Kantorovich, S.Knyazev, A.Shrubenko, etc.

In spite of the obtained success there are many unsolved present-day problems. In particular, insufficiently studied are mechanisms for representing heterogeneous innovative knowledge in a manner that ensures its rapid development and versatile use. Traditional models (production, semantic, etc.) do not solve the problems under consideration, because the models are narrow-focused, complex and difficult for decomposing into components. Furthermore, there are neither interactive models ensuring synchronization of activities between sources and consumers of innovations nor mechanisms for rapid accumulation of such knowledge and making the corresponding decisions, the relevance of which is confirmed by the competence of experts and compliance with the environment requirements. There is also a lack of easy-to-use and inexpensive technologies to construct systems ensuring the acquisition and mainstream use of the innovative knowledge.

The dissertation attempts to comprehensively investigate and solve the stated problems on the basis of the synthesis of the decision-making theory, organization theory, artificial intelligence and possibilities of modern software technologies.
OVERVIEW OF THE DISSERTATION STUDY

The dissertation and major research programs and themes

The results of the dissertation are obtained through research carried out at the Department of Information Management Systems, Faculty of Applied Mathematics and Computer Science, Belarusian State University (BSU), in accordance with:

1) scientific and research programs of BSU:
   – “Development of General Principles for Constructing Intelligent Information Systems, Mathematical Models, Methods, Information and Computer Technologies for Decision Making and Pattern Recognition Systems”: a task within the scientific and research program of BSU for 2006-2010;
   – “Development of Theoretical Foundations and Technologies for Constructing Information Management Systems with the Application in Intelligent Data Mining”: a task within the scientific and research program of BSU for 2011-2015;

2) scientific and technical programs of the Republic of Belarus:

3) grants from the Belarusian Republican Fund for Fundamental Research (BRFFR) of the National Academy of Sciences of Belarus:
   – “Development of the Algebraic Theory of Recognition and its Application in Intelligent Data Mining”: a task № F10P-097 being performed within the decision of BRFFR Council for 2010-2012.

The purpose and objectives of the research

The purpose of the dissertation research is the development of models, algorithms and the technology for decision making intellectualization on the basis of distributed innovative knowledge.

To achieve the purpose the following objectives are stated and realized:

1. To analyze modern methods, information and software technologies for decision making intellectualization.

2. To develop a meta-model, which integrates heterogeneous knowledge represented in various formats and structures, and an environment model that ensures synchronization of activities between the distributed sources of innovations and consumers of the meta-model content.
3. To develop: 1) an algorithm for constructing and updating the meta-model content, the algorithm should be invariant to the number of sources; 2) an algorithm for constructing the scene ensuring the life cycle of the content; 3) the corresponding decision-making algorithm ensuring the choice of the required variant.

4. To develop a computer technology for minimizing time and expenses for building, implementation and distribution of the meta-content.

The object of the research: computer systems for decision support.

The subject of the research: models and algorithms for making decisions on the basis of distributed expert knowledge.

Claims of the dissertation

1. The meta-model for representing heterogeneous knowledge in the form of subject collections and the corresponding object model, thus ensuring a solution to the problem of structural, semantic and language interoperability of heterogeneous models of knowledge.

2. The information-technological model of an open interactive scene for information interaction between distributed actors on the basis of cloud deployment of software, information and computing resources, thus ensuring formation, use and updating of subject collections within the global environment.

3. The algorithms for constructing and updating subject collections of distributed experts, thus ensuring the constant improvement of the content, as a result, the problem of knowledge aging is largely solved and the time for innovation implementation is reduced.

4. The algorithm for decision making based on subject collections. The algorithm provides a solution from a number of alternatives with regard to independent evaluation and experience of content usage, thus ensuring the result, which is adequate to the requirements of the user and the environment, as well as improvement of the content through the use of external experience.

5. The distinctive technology for building computer systems to support decisions on the basis of subject collections with the use of the cloud approach, thus reducing the expenses for the development of the system and the support of the life cycle of software and hardware.

Personal contribution of the author

The dissertation is based on the results of the author's scientific investigation carried out at the Department of Information Management Systems, BSU. The scientific adviser participated in the problem statement and discussions of approaches to
the problem solution. In joint publications the results were obtained personally by the author of the dissertation.

**Approbation of the results**

The main scientific and practical results of the dissertation were reported and discussed at the following international and Belarus Republic conferences:


**Publication of the dissertation results**

The findings of the dissertation were published in 26 printed works, including: 8 articles in reviewed scientific journals (2.8 author’s sheets) in accordance with item 18 of the “Statute on the Awarding of Academic Degrees and Conferring Academic Ranks in the Republic of Belarus”, an article in the reviewed scientific journal “Tavricheski Vestnik of Informatics and Mathematics“, 17 articles in proceedings of scientific conferences.

**The structure and volume of the dissertation**

The dissertation consists of the list of abbreviations, the introduction, the overview of the dissertation study, 4 chapters, the conclusion, the bibliography and 2 appendices. The total volume of the dissertation is 104 pages, including 37 illustrations on 10 pages, 4 tables on 2 pages, 2 appendices on 8 pages, the list of 162 references comprising printed works of the author on 11 pages.

**THE MAIN CONTENT OF THE DISSERTATION**

Chapter 1 considers main problems of decision making in the new information
and communication environment. An ontology-based approach (IDEF5) is used for presenting the results of the research. The approach offers the possibility to describe the solution of complex problems from the time of the conceptual model construction till obtaining the corresponding program code.

In general, a Decision Making Problem (DMP) is reduced to the choice of one or more alternatives from a variety of options for resolving the problem situation. DMP is described by the tuple:

$$DM = Alg (S, G, E, V, Kr, V)$$  \hspace{1cm} (1)

where: $S$ signifies problem situation; $G$ stands for the goal; $V$ signifies possible alternative solutions; $Kr$ is a selection criterion; $Alg$ is an algorithm; $V$ stands for the solution result. A typical solution scene is described by the tuple:

$$Scene = (C, E, U, DM, DSS, com)$$  \hspace{1cm} (2)

where: $C$ is a decision initiator (the Center); $E$ stands for information sources; $U$ signifies users of the decision; $DM$ is a decision making problem; $DSS$ stands for a decision support system; $com$ denotes means for information exchange between $C$, $E$, $U$.

Traditionally, the synthesis of the DSS-based solution is carried out for tasks with long life cycle within local scenes on the basis of permanent sources of knowledge and rigid schemes of communication. Expert knowledge or the results of the statistical analysis of time series, formalized within logical, production or other models, are used to form DMP constituents. The number of users of the decision is limited, as a rule, by the Center ($C$). DSS knowledge base is characterized by the difficulty of decomposition into components and implementation, resulting in narrow-focused use and rapid knowledge aging. Synthesis of the solution $V$ generally requires large investment of time and money. As a result, DSS users are mainly government agencies and large companies.

Properties of the new environment and peculiarities of decision making are determined on the basis of the analysis of the existing studies. It is shown that as a result of globalization, properties of DMP (1) and the scene of its solution (2) are significantly changed, in particular:

- knowledge has acquired the status of the primary means for obtaining competitive advantages. This substantially increases the role of technology-oriented innovative knowledge represented in various models and formats, offering the possibility of fragmentation, export, and versatile use. As a result, there is a problem of compatibility (interoperability) between homogeneous models (logical, production, etc.) and heterogeneous cognitive structures;

- globalization of the environment has changed the scene of the solution (2), there is a need for remote communication between distributed actors for creation, updating and use of innovative knowledge bases;
increase in number of users has led to the need for the development of inexpensive, easy-to-use DSS ensuring the remote access to innovations, construction of DMP constituents, and the choice of effective solutions.

It is also shown, that traditional methods and technologies can not overcome the emerged contradictions, therefore, a number of scientific schools are trying to solve them within the so-called “problem of intellectualization”. This term refers to a wide range of problems for developing protective mechanisms to support the homeostasis of natural and artificial systems on the basis of competent knowledge and objective evaluation.

A new approach to the intellectualization of decision making is proposed. The approach is based on subject collections as a form of representing heterogeneous knowledge, the relevance of which is confirmed by the competence of sources and external (independent) estimates.

A task for the development of the computer technology for decision making intellectualization is stated. The technology should include the construction of: 1) a meta-model for heterogeneous knowledge representation; 2) an open environment model for implementing its life cycle; 3) algorithms and software for the creation of the corresponding DSS.

Chapter 2 addresses theoretical problems of decision making intellectualization in the global environment. The problems concern representation of innovative knowledge in the form convenient for multifunctional application, printing, replication and export; construction of mechanisms for knowledge acquisition and permanent updating; decision making based on distributed innovative knowledge.

It is shown that the existing models (productions, semantic networks, frames, patterns, etc.) are homogeneous in nature and technologically are not compatible with the heterogeneous nature of innovative knowledge.

The concept of a subject collection (SC) is proposed as a solution to the problem of interoperability. A subject collection is an abstract structure that integrates heterogeneous models of knowledge as components of the general solution. A distinctive feature of the concept of SC is an inextricable connection between the notion of knowledge and mechanisms of its synthesis and application, thus ensuring objective assessment of the utility level.

A system of axioms (the existence, representation, heterogeneity, innovation, relativity and applicability) is formulated. The axioms make it possible to determine the limits of the use and life cycle of a SC. Based on the proposed axioms, a graph model of the SC is developed. The model ensures identification of the SC in the global environment taking into account the roles of all participants (Figure. 1).
The vertices of the graph identify the SC in the global (IdG), subject (IgD) and corporate (IdC) environment and determine actors, i.e. the initiator (C) of the SC formation, sources of knowledge (E) and users (P). The solution of task Z is represented by a set of algorithms (Alg), technologies (Tech) and experience (Exp), the usefulness of which are confirmed by the external evaluation. Independent suggestions for improving knowledge content are represented in $\Delta_1^{\ldots}\Delta^n$.

For practical use of SCs an object-oriented version is developed, thus ensuring implementation of SC meta-model in any modern programming language. This solves the problem of semantic interoperability of heterogeneous models of knowledge models. A typical XML is proposed for SC storage, thus solving the problem of integration of the content and its elements into external software systems.

Next, the problem of scene modeling for the implementation of SC life cycle is considered. The universal model of the scene participant (actor) is proposed:

Actor = (address, Name, Place, Status, inf, Role, Dlg (Q,R)),

where: address is actor address in the network; Name is a name; Place stands for location; Status denotes status; inf signifies additional information; Role is a role; Dlg signifies a dialogue (Q is a question, R is an answer). The model forms the basis for the construction of models of C, E, U actors.

It is shown that the existing scene models do not ensure support of SC life cycle. It is proposed to add participants of social networks to the scene (for objective evaluation of SC usefulness) and an artificial actor (“Validator”) to control access to SCs (Figure 2).
As a result, the obtained scene ensures remote access to SCs, objective assessment of the content and a controlled increase in the number of users by participants of social networks. The proposed scene construction algorithm is based on the integration of standard models of actors into an open scene with feedback.

**Algorithm 1.** The input elements are details about C, E, UG, Twitter, Facebook actors. The output is a scene formalized in the database of actors (Figure 3).
The algorithm forms a distributed open scene, including social networks as a source of objective evaluation of content usefulness, and potential SC users. The formed scene is a tool for objective decision making and the basis for the construction of the mechanism for protecting various categories of users from outdated or false solutions in the web.

After the scene construction the problem of SC content formation is considered. It is shown that the existing algorithms do not solve this problem, because they are focused on non-decomposable, difficult to change and expensive knowledge ba-
ses. The algorithms also lack mechanisms for independent content assessment and considerations of its improvement.

The algorithm, based on "Lego" principle, is proposed. Separate fragments, in total, form an entire view. First, the Center C, through the dialogue dRoleC, forms SC as a template, which includes the classification features and the description of the problem indG, ..., Z, the remaining fields are left blank. Then, an expert (within the dialogue dRoleE) forms the SC content adequate to the stated problem and sends the result to the Center.

Algorithm 2. The input element is a task. An expert decomposes it into sub-tasks and forms the theoretical and technological solution. The output elements are the constructed fragments of the subject collection: nZ, Z, Alg, Tech, Exp(Figure 4).

Figure 4 – The flowchart of the algorithm for the subject collection formation
As a result, a heterogeneous subject collection is formed, which contains formalized expert knowledge, necessary and sufficient for the theoretical and practical solution. Unlike the traditional non-decomposable Lisp-Prolog databases, any fragment of the content is considered as a logically and physically independent unit that can be rejected, exported and used in external software systems.

The second part of Chapter 2 considers problems of decision making on the basis of SC. It is shown that classical methods of decision making usually focus on the closed environment and the narrow range of tasks. They lack mechanisms for independent evaluation of the results and adaptation to environment changes, thus making it difficult to intellectualize the process of decision making. Construction of solutions in an open environment has its own peculiarities: alternatives are formed by experts in various centers of innovation; the number of alternatives often reaches several thousand; the method of choice depends on the purpose, user competence, etc.

In view of SC structure and open interactive environment (OIE) composition, the general problem of decision making (1) can be described by the tuple:

$$\text{DM} = \text{Alg}(Z, f, \Omega, \text{OIE}, SC_1, SC_2, ..., SC_k, SC^*)$$ (3)

where: $Z$ is a problem identifier, $f$ is a task type; $\Omega$ signifies the number of chosen solutions (threshold), corresponding to $f$; OIE is an open interactive environment; $SC_1, SC_2, ..., SC_k$ are subject collections; $SC^*$ signifies the chosen solution.

A new approach to the solution of the problem (3) is proposed. The approach is based on the synthesis of the elements of the utility theory (F.Edgeworth, V.Pareto, I.Fisher) for the numerical evaluation of content quality, and the principle of equilibrium (J.Nash) for choosing a solution, which is adequate to user requirements and the environment state. To achieve consistency between user’s purposes and the result of the choice it is proposed to divide all decision making problems into two types. The first type concerns mass problems ($f_1$), requiring the solution with maximum utility assessment. The second type are professional problems (projects) ($f_2$), using solutions balanced with the environment requirements.

In view of the peculiarities of the problem (3), two levels are considered for obtaining a solution: construction (search and narrowing) of a set of subject collections $SC_1, SC_2, ..., SC_k$, belonging to the given problem; the choice of the solution satisfying the user requirements. At the first level a set of admissible solutions is formed. Initially, a set of all subject collections (available alternatives) is determined in accordance with the problem identifier $Z$. Then, on the basis of the analysis of estimates (ind), this set is narrowed to a subset with the proven level of content usefulness. For the given threshold $\Omega$, alternatives with maximum utility indices are selected as candidates for the solution.

At the second level the solution is chosen. Depending on the type of the problem to be solved, the choice is made in accordance with the level of utility (for type
f1) or on the basis of equilibrium principle (for type f2). In the former case, the solution with the maximum utility index is chosen. In the latter case, the utility of each alternative in practice is considered, accordingly, the value of index (ind) is adjusted and the equilibrium solution, which meets the requirements of the user and the environment, is chosen.

**Algorithm 3.** The input element is identifier of problem nZ. The output elements are the fragments of the SC: nZ, Z, Alg, Tech, Exp, ind, Δ (Figure 5).

![Flowchart](image)

**Figure 5 – The flowchart of the two-level algorithm for decision making**

The principle of the guaranteed result ensures the choice of the solution with the maximum index of usefulness. The solution can be considered as the best one for beginners. According to the principle of equilibrium, a set of solutions for the given threshold (with the decrease of the index value) is generated. A compromise solu-
tion, which meets preferences of the user and the environment, is chosen from the set. In practice, this means the possibility of a consensus between capabilities of the developer and customer's requirements.

The evaluation of SC content usefulness can be considered as objective, since the evaluation is made by target groups and independent members of social networks. Information about content usage is recorded in fields of utility evaluation and suggestions for the improvement, serving as the basis for the feedback "expert-user-expert." Lowered rating of content usefulness by a large number of users is a signal for the Center and experts to begin content updating. The corresponding flowchart of the algorithm for content updating is shown in Figure 6.

**Algorithm 4.** The input elements are: \( \text{ind} \neq 0, \Delta \neq \emptyset, \text{Alg}, \text{Tech}, \text{Exp} \). The output is an updated subject collection: \( \text{Alg}', \text{Tech}', \text{Exp}' \), \( \text{ind} = 0, \Delta = \emptyset \).

![Figure 6 – The flowchart of the algorithm for content updating](image-url)
Thus, the proposed models and algorithms ensure the full-scale solution to the problem of intellectualization by the use of innovative expertise.

**Chapter 3** is devoted to the development of software architecture for automating the processes of decision making on the basis of SCs. In accordance with the purpose of the dissertation, the architecture must ensure the functioning of the processes in a distributed environment with the use of various devices (PCs, smartphones, tablets, etc.). It is shown that typical DSS architectures mainly use a local paradigm and do not ensure the fulfillment of the above-mentioned requirements. An architecture is proposed that is based on the replacement of a local paradigm with an open one. As a result, there is a possibility of the distributed processing of the content and unlimited increase of SC users. A multi-agent approach is used for implementing the methodology. The approach is originally designed to solve distributed problems. The corresponding architecture is proposed (Figure 7).

![Multi-agent architecture](image-url)
The architecture includes the Center, Expert, User agents, implementing basic processes Pr1 (SC initiation), Pr2 (content creation and updating), Pr3 (decision making and evaluation of the content usefulness); as well as the Validator agent to regulate access (Pr4) to resources. All agents are built on the single basic scheme <sensor-effector-processor-memory>. Specific features of their tasks in a distributed environment are taken into account. The use of the universal XML-database of subject collections and the implementation of dialogues via a Web browser facilitate the integration of the multi-agent architecture into a variety of external environment including a cloud one.

The autonomy of agents ensures the homeostasis of the architecture within changes of environment requirements. The integration of the architecture into the cloud structure offers the possibility of unlimited increase in the number and volume of SCs and improves the content security. The use of interfaces of the cloud provider ensures access to SCs from different types of devices (PCs, smartphones, tablets), which contributes to the increase in the number of content users.

Chapter 4 describes the composition of software and methods of its use for making decisions on the basis of subject collections.

Capabilities of modern software and cloud technologies, needed to implement the architecture, have been analyzed. It is shown that an effective tool for SC formation is MS.NET software platform, which provides synthesis of the executive code based on different languages (Prolog, C++, C#, etc.). The scheme of content formation with the use of .NET in combination with other means is shown in Figure 8.

**Figure 8 – The scheme of SC formation**

It is proposed to use Drupal’s open source for building a private cloud resource. The source includes the Apache server and PHP language. The portal “Subject
Collection” is designed and hosted in the cloud resource of Byelex company on the basis of the chosen software (Figure 9).

The technique of using the portal for practical applications is developed. The access control, formation, updating, use and evaluation of subject collections are described.

![Figure 9 – SC export to external media](image)

The final part of the chapter provides examples of the developed, within the dissertation, local and network SCs for decision making support in the field of medicine: “Orthopedishe Casuistiekk“ and “Ortho-Expert“ (for orthopedists); “Atlas van de Parodontale Diagnostic“ and “Atlas Mond- & Kaakziekten“ (for dentists); “Atlas of Forensic Medicine“ (for forensic medical examination).

Thus, the use of MS.NET for constructing SC content ensures the solution to the problem of interoperability of knowledge represented in different languages. The integration of the private portal resource into the cloud one makes it possible to reduce expenses for the purchase and support of the life cycle of OS, software and hardware due to the inheritance of constantly updated resources of the provider.
CONCLUSION

Main scientific results of the dissertation

1. The meta-model for representing heterogeneous knowledge in the form of a subject collection and the corresponding object model, ensuring the solution to the problem of structural, semantic and language interoperability of heterogeneous models of knowledge represented in different formats, are proposed. The use of the object-oriented approach makes it possible to represent rigid heterogeneous models as one class of methods, consistent performance of which solves the general problem. As a result, there is a possibility to combine previously developed and new methods, inherit the results, to export meta-model elements and update any method without changing other ones [1, 2, 9, 17, 18].

2. The information-technological model of the open interactive scene (organizational structure) for the information exchange between distributed actors through cloud deployment of program, information and computing resources is developed. The model offers the possibility to form, use and update subject collections in the global environment [3, 4, 8, 14, 16, 20, 21, 22].

3. Algorithms for constructing and updating subject collections of distributed experts are developed, thus ensuring the permanent improvement of the content. As a result, the problem of knowledge aging is largely solved and the time for innovation implementation is reduced [8, 9, 12, 13].

4. The decision-making algorithm, based on subject collections, is developed. The algorithm offers the possibility to choose the solution from a number of alternatives taking into account the independent evaluation and experience of the content usage. This ensured the result adequate to the requirements of the user and the environment as well as improvement of the content through the use of the external experience [5, 7, 11, 15, 19, 23, 25].

5. The distinctive technology for building computer systems to support decision making on the basis of subject collections with the use of the cloud approach is developed. The technology and advantages of the environment model ensured considerable reduction of expenses for the development of the system and the support of the life cycle of software and hardware [6, 8, 10, 12, 24, 25, 26].

Recommendations for the practical use of the results

Theoretical and technological results, obtained in the dissertation, can be used for building computer systems based on innovative knowledge, for supporting decision making in the rapidly developing fields including the computer technology, medicine, education, ecology, etc.
PUBLICATIONS RELATING TO THE THEME OF THE DISSERTATION

Articles in scientific journals


Articles in proceedings of scientific conferences


20. Краснопрошин, В.В. Принятие решений в оперативных задачах регионального управления / В.В. Краснопрошин, Х.Виссия, А.Н. Вальвачев // Труды


SUMMARY

Vissia Herman Elisabeth Rene Maria

MODELS, ALGORITHMS AND THE TECHNOLOGY FOR DECISION MAKING INTELLECTUALIZATION BASED ON SUBJECT COLLECTIONS

**Keywords:** decision making, intellectualization, knowledge engineering, innovations.

**The purpose of the research:** development of models, algorithms and the technology for decision making intellectualization based on innovative knowledge.

**Research methods:** decision making theory, organization theory, knowledge engineering, multi-agent approach, software technologies.

**The obtained results and scientific novelty.** New models, algorithms and the technology for the decision making intellectualization are developed. The concept and model of subject collections are proposed. The model, based on the theory of patterns, ensures representation of heterogeneous expert knowledge. The scene model is developed. The model, based on the theory of open systems, ensures formation, objective assessment and updating of subject collections in the global environment.

The following algorithms are constructed: the algorithm for forming and updating the content of a subject collection, the algorithm is invariant to the number of knowledge sources; the algorithm for constructing the scene that ensures the life cycle of the content; the decision-making algorithm that ensures the choice of an alternative with regard to independent evaluations and experience of content usage.

Implementation of models and algorithms in the form of a portal on the basis of .Net, CMS platforms and cloud technologies offers the possibility to reduce costs for developing the decision making system and for supporting the life cycle of software and hardware.

The obtained theoretical and technological results allow the developer to construct low-cost, easy-to-use systems that ensure rapid accumulation and mass use of innovative knowledge.

**Usage recommendations.** The developed software can be used for building computer systems that use innovative knowledge to support decision making in the rapidly growing areas of the economy, such as the computer technology, medicine, education, ecology, etc.