International conference

INFORMATION RESEARCH, APPLICATIONS AND EDUCATION

14-24 June 2004, Varna

i.tech

PROCEEDINGS

FOI-COMMERCE

SOFIA, 2004

Kr. Markov (Ed.)

Proceedings of the Second International Conference “Information Research, Applications and Education” i.TECH 2004, Varna, Bulgaria

Sofia, FOI-COMMERCE – 2004

ISBN: 954-16-0031-X
First edition

Printed in Bulgaria by FOI-COMMERCE
Sofia -1090, P.O. Box 775
e-mail: foi@nlcv.net
www.foibg.com

ISBN 954-16-0031-X
C\o Jusautor, Sofia, 2004

Preface

The International Conference "Information Research, Applications and Education" (i.TECH) is organized as a part of "ITA 2004 - Joint International Scientific Events on Information Theories and Applications".

The main organizer of the ITA 2004 as well as the i.TECH 2004 is the

International Journal on Information Theories and Applications (IJ ITA).

The aim of the conference is to be one more possibility for contacts for IJ ITA authors. The usual practice of IJ ITA is to support several conferences at which the IJ ITA papers may be discussed before submitting them for referring and publishing in the journal. 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.

I would like to express my thanks to all who support the idea to organize the i.TECH 2004.

Let me thank the Program Committee of the conference for referring the submitted papers. Special thanks to Mr. Petar Barnev and Mr. Avram Eskenazi – Co-chairs of the Program Committee.

i.TECH 2004 Proceedings has been edited in the Institute of Information Theories and Applications FOI ITHEA and published by FOI COMMERCE Co.

The i.TECH 2004 Conference found the best technical support in the work of the Technical editor Ms. Krassimira Ivanova and Organizing secretary Mr. Ilia Mitov.

To all participants of i.TECH 2004 I 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.

Varna, June 2004

Krassimir Markov

IJ ITA Editor in Chief

i.TECH 2004 has been organized by:

IJ ITA - International Journal on Information Theories and Applications
Institute of Information Theories and Applications FOI ITHEA
Institute of Mathematics and Informatics of BAS
Association for Development of the Information Society

The main topics of the i.TECH 2004 include but are not limited to:

Business Informatics
Extreme programming
Education informatics
Information systems
Information Modelling
Quality of the programs
Multimedia systems
Applied program systems
Software engineering
Statistical systems
Hyper technologies

Official languages of the conference are Bulgarian, Russian and English.

General sponsor of the i.TECH 2004 is FOI BULGARIA (www.foibg.com).

TABLE OF CONTENTS

PREFACE ... 3

INTAS-FET Strategic Workshop

DATA FLOW SYSTEMS: ALGORITHMS AND COMPLEXITY ... 9

FOURIER NEURAL NETWORKS: AN APPROACH WITH SINUSOIDAL ACTIVATION FUNCTIONS
L.F de Mingo, L. Aslanyan, J. Castellanos, M. A. Díaz, V. Riazanov ... 14

MULTI-AGENT SYSTEMS IN THE HARVEST PROGNOSIS
Laura Ciocoiu, Cristian Paraschiv, Barbu Dragoş ... 16

Papers in English

EXPERIENCE WITH WEB-BASED EDUCATIONAL PLATFORM – POSSIBILITIES AND LIMITATIONS
Petya Assenova, Georgy Todorov ... 21

PROJECT SOLUTIONS REUSE
M. Bondarenko, A. Kuzemin, M. Sorochan, I. Yanchevskiy ... 27

REPRESENTING "RECURSIVE" DEFAULT LOGIC IN MODAL LOGIC
Frank M. Brown ... 28

ON THE RELATIONSHIPS AMONG QUANTIFIED AUTOEPISTEMIC LOGIC, ITS KERNEL, AND QUANTIFIED REFLECTIVE LOGIC
Frank M. Brown ... 37

METHODS FOR SOLVING NECESSARY EQUIVALENCES
Frank M. Brown ... 45

ABOUT NEW PATTERN RECOGNITION METHOD FOR THE UNIVERSAL PROGRAM SYSTEM “RECOGNITION”
A.A. Dokukin, O.V. Senko ... 54

TEACHING STRATEGIES AND ONTOLOGIES FOR E-LEARNING
Tatiana Gavrilova, Michael Kurochkin, Victor Veremiev ... 59

QUALITY AND EFFICIENCY IMPROVEMENT OF THE TEACHING PROCESS OF MATHEMATICS BY APPLICATION OF SPECIALIZED SOFTWARE
Tsvetanka Kovacheva ... 66

MUSIC AS THE SOURCE OF INFORMATION INFLUENCE AND SOUL EDUCATION
L. Kuzemina ... 69

FORMAL DEFINITION OF THE CONCEPT “INFOS”
K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova ... 71

A TWO LAYERED MODEL FOR EVOLVING WEB RESOURCES
A. Milani, S.Suriani ... 75

XML EDITORS OVERVIEW AND THE CHALLENGE TO BUILD XML-ORIENTED EDITOR FOR MEDIAEVAL MANUSCRIPT DESCRIPTIONS
P. Pavlov ... 80

LOCAL GOALS DRIVEN HIERARCHICAL REINFORCEMENT LEARNING
A. Pchelkin ... 88

SITUATION CENTERS IN MODERN STATE
M. Slipchenko, A. Kuzemin, M. Sorochan, A. Torojev ... 96

AN INTERACTIVE METHOD OF LINEAR MIXED INTEGER MULTICRITERIA OPTIMIZATION
V. Vassilev, K. Genova, M. Vassileva, S. Narula....................99

DIGITAL CREATIVITY: ADVANTAGES, PROBLEMS, RESPONSIBILITIES
V. Velev....................105

XML PRESENTATION OF DOCUMENTS USED FOR DATA EXCHANGE IN MANAGEMENT OF GENETIC RESOURCES
L. Yordanova, Vl. Dimitrov....................113

ALGORITHM BIDIMS FOR AUTOMATED SYSTEMATIZATION OF DATA ARRAY. CASE STUDY: REDISCOVERING MENDELEEV'S PERIODIC TABLE OF CHEMICAL ELEMENTS.
Andrey N. Zagoruiko, Nickolay G. Zagoruiko....................119

Papers in Russian and Bulgarian

ПРИМЕНЕНИЕ НЕКЛАССИЧЕСКИХ ЛОГИЧЕСКИХ МЕТОДОВ В ЗАДАЧАХ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ И ПОИСКА ЗАКОНОМЕРНОСТЕЙ В БАЗАХ ЗНАНИЙ
В.С. Йоцов, В.С. Сгурев, А.В. Тимофеев, Л.В. Лютикова....................131

АДАПТИВНОЕ УПРАВЛЕНИЕ И САМООРГАНИЗАЦИЯ В МУЛЬТИ-АГЕНТНЫХ ИНФОКОММУНИКАЦИОННЫХ СЕТЯХ
А.В.Тимофеев....................139

ЕДНА РЕАЛИЗАЦИЯ НА ХЕШ-ТАБЛИЦА С ОТВОРЕНО АДРЕСИРАНЕ В ПОСЛЕДОВАТЕЛНО РАЗПРЕДЕЛЕНА ПАМЕТ
В. Дянкова, Р. Христова....................144

ОРГАНИЗАЦИЯ НА ДИАЛОГА С ПОТРЕБИТЕЛЯ В КОМПЛЕКС ФОИ
Красимира Иванова, Илия Митов, Красимир Марков....................150

РАЗШИРЕНА ИНФОРМАЦИОННА СИСТЕМА ЗА РЪКОВОДИТЕЛЯ (EXTENDED EIS)
Тодорка Ковачева....................156

КООРДИНАТНО БАЗИРАНА ФИЗИЧЕСКА ОРГАНИЗАЦИЯ ЗА КОМПЮТЪРНО ПРЕДСТАВЯНЕ НА ИНФОРМАЦИОННИ ПРОСТРАНСТВА
Красимир Марков....................163

ЕДИН ПОДХОД ПРИ АНАЛИЗ НА ПРЕДПОСТАВКИ ЗА ИЗГРАЖДАНЕ НА Е-ОБЩИНА
Димитър Николов, Румяна Стефанова, Ивелин Стефанов....................173

ИНФОРМАЦИОННО МОДЕЛИРАНЕ НА ДВУКООРДИНАТНО ИЗМЕРВАНЕ НА ОПТИЧНИ ПАРАМЕТРИ
Георги Стоилов, Петър Шарланджиев, Никола Мечкаров....................184

INTAS-FET STRATEGIC WORKSHOP

“Data Flow Systems: Algorithms and Complexity”

DATA FLOW SYSTEMS: ALGORITHMS AND COMPLEXITY[1]

Projects:

– INTAS 00-397 “Data Mining Technologies and Image Processing: Theory and Application”
– INTAS 00-626 “Data mining algorithm incubator"

Project representatives:
J. Castellanos, INTAS 00-626 Coordinator and *F. Mingo*, Technical University of Madrid
M. Mulvenna, University of Ulster
K. Vanhoof, Limburg University
M. Hatzopoulos, University of Athens
L. Aslanyan, Workshop Co-chair, Institute for Informatics and Automation Problems of Armenian National Academy of Sciences
K. Markov, Workshop Co-chair, Institute of Information Theories and Applications
V. Ryazanov, Computer Center of Russian Academy of Sciences
S. Ablameyko and *A. Tuzikov*, United Institute for Information Technologies of Belarusian Academy of Sciences
G. Katona, A. Renyi Institute for Mathematics of Hungarian Academy of Sciences
H.D. Gronau, Rostock University
J. Vorachek, INTAS 00-397 Coordinator, Lappeenranta University of Technology
A. Ionita, Research Institute for Artificial Intelligence, Romanian Academy of Sciences
L. Ciocoui, National Institute for Research and Development in Informatics
Yu. Zhuravlev, Moscow State University
H. Sahakyan, Information Society Technologies Center Armenia
V. Gladun, V.M.Glushkov Institute of Cybernetics of Ukrainian Academy of Sciences

Abstract

Data Flow Systems: Algorithms and Complexity (DFS-AC) - names the current research objective of a cluster of international RTD projects[2]. It brings together not only research directions such as algorithms and complexities, but also machine intelligence and distributed systems, towards the integrated research of principles of novel information systems, – the so called framework of Hereditary Hybrid Societies (HHS). Algorithmic and complexity concepts are investigated regarding the novel issues of extremely high and extending size data amounts and data flows in networks. Mathematically - the main emphasis is in a novel combination of application oriented researches in three important areas - massive data sets, autonomous agents - living and communicating in a society, and complex systems. Technological component is in knowledge acquisition and accumulation through distributed and co-operating systems, in study of properties, approximating intelligence in nature and in human society.

New Initiative

An open online consultation forum has been recently started (HTTP://WWW.CORDIS.LU/IST/FET/ID.HTM) at IST FET (Future and Emerging Technologies) of 6th FP Research Programme of European Commission, to allow submitting and improving ideas for new promising IST-related research areas which might become future FET proactive initiatives to be funded by the EU through future IST calls for proposals.

The two “new initiative” submissions, presented below, - were prepared by the INTAS 00-397 and INTAS 00-626 teams and the INTAS-FET Strategic Workshop “Data Flow Systems: Algorithms and Complexity” aims at larger discussions of these issues in framework of the International Conference "i.TECH 2004".

---

[1] Research partially supported by INTAS, http://www.intas.be.

[2] NATO HTECH-LG-961170 "Distributed Information Systems for Critical Social Issues"; INTAS 96-952 "Concurrent heuristics in data analysis and prediction"; INTAS 00-397 "Data Mining Technologies and Image Processing: Theory and Application", and INTAS 00-626 "Data Mining Algorithm Incubator".

Title of new initiative 62:

Introspective Reasoning

http://fp6.cordis.lu/ist/fet/proposal_details.cfm?ID_EVENT=62

Research Objectives and Challenges

Predicting the future is a difficult task. But that is the challenge that has been taken up by areas of artificial intelligence and in particular data mining. At this time, in 2004, we have had relative success in well-defined areas; for example, online recommender systems and discrete scientific domains. But the predictive systems developed today are brittle in that they cannot deal with major changes in data sources or trends across time and other dimensions.

Set against this landscape, our world is becoming increasingly a digital one. The amount of data collected is increasing as sensor technology evolves, and as more people use the Internet and its attendant services.

Consequently we need to address the challenge of building flexible reasoning systems. Such systems must be open source, non-brittle reasoning systems that adapt to new environments and their data, that detect changes and trends, and that can measure their own performance.

This requires research that addresses several interconnected areas:

1) Concept drift. Detecting drift in data sources.
2) Ensemble reasoning/recommendation. Bringing together different predictive models. Based on distributed, disparate data sources or indeed knowledge base. Using different mediation schemes between the recommendation and predictive agents.
3) Introspection. Critical appraisal of the predictions/recommendations to improve knowledge structures.
4) Agent-based computing and web services. Required to facilitate distributed, heterogeneous systems.
5) Semantic-based technologies. Ontologies.
6) Representation. Using and extending PMML (www.pmml.org), RDF, OWL, DAML+OIL
7) Extending Open Standards such as data mining APIs like JSR-73 (http://www.jcp.org/en/jsr/detail?id=73)

Rationale

The constituent research areas of Introspective Reasoning cover a broad spectrum of computer science, but critically combine together several strands of research that are novel and high risk. Importantly the introspective capability provides a means of objectively measuring the performance of implemented systems. For example for online recommenders, the success of personalisation can be measured, and serve as an indicator of user acceptability to online services.

The European Commission funds many RTD projects that have a component which is data mining, recommendation, personalisation, etc. Much of this research funding is not being used effectively. What is required is a research drive towards standards-based, open source, extensible predictive software that offers measurable performance and non-brittle adaptability.

Past projects have also tried to build a theory of what predictive techniques are more relevant under what circumstances. It is now accepted that such automated techniques for choosing the most appropriate predictive model is not achievable and more context sensitive approaches to learn dynamically within an ensemble (circumstances when one model is more appropriate than others) is required. It is timely that many of the strands of research required for introspective reasoning are maturing (e.g., agent technology and ensemble reasoning) or offer significant promise (concept drift, ensemble recommendation). There is a synergistic potential across the research centres in Europe to push forward in Introspective Reasoning, especially as much of the global expertise and knowledge base for this research is based in our continent. It is not possible for single research organisations in Europe to address such a challenge on their own, which is why support and stake-holding from the European Commission is required.

Impact

The focus of IST in FP6 is on "the future generation of technologies, in which computers and networks will be integrated into the everyday environment, rendering accessible a multitude of services and applications through easy-to-use human interfaces. This vision of ambient intelligence places the user, the individual, at the centre of future developments for an inclusive knowledge-based society for all."

We are in danger of becoming a data-rich, knowledge-poor society. If the trends continue, then key objectives of the Information Society will be jeopardised. Social exclusion to e-services will be the norm, as the percentage of digitally disadvantaged increases as Europe's median age increases. What is required to power the visions of 'ambient intelligence' and 'easy to use human interfaces' is standards-based, open source, extensible predictive software that offers measurable performance and non-brittle adaptability.

The potential impact on Europe at societal, economic and business level is significant. An increasing amount of our disposable income is now being used for online services, while essential public services (from government to health to education etc) are moving towards new modes of support and delivery. This will bring about a change in business models, with attendant societal impact. Providing Introspective Reasoning adds value to the social provision model, and minimises the potential for an increase in the digital divide.

Giving Europe a lead in this area will lead to an increase in spin-off and joint ventures generating fast-moving businesses that help support the coming transition in e-services in Europe.

Communities addressed, other related initiatives

AI, Cognitive science, behavioural psychology, agent based research, data mining and predictive modelling, semantic web, Grid.

Contact Details

Mr Maurice Mulvenna from University of Ulster; E-mail: md.mulvenna@ulster.ac.uk

Title of new initiative 69:

Hereditary Hybrid Societies (HHS)

http://fp6.cordis.lu/ist/fet/proposal_details.cfm?ID_EVENT=69

Research Objectives and Challenges

HHS research area is originally aimed to study the highest intellectual properties of advanced information ambience in similarity with the behaviour of nature and societies. The main idea presented to develop and implement appropriate “modelling” and “algorithmic” solutions for the globally distributed autonomous activities, huge data amounts and data flows dynamically increasing in size, knowledge communication and accumulation. It is challenging to gain the IT Inheritance (accumulated transfer of knowledge, properties, etc.), originally a sophisticated property of nature.

Today’s paradigm is that Internet is breaking the traditional ways of thinking. It deals with heterogeneous dynamic resources through interaction of mobile and intelligent agents (both machine and human) creating a hybrid society with more complex sophisticated properties, approaching inheritance. A broad multidisciplinary research of behavioural knowledge models and novel mathematics is necessary to create hybrid societies, which is the internals of advancing AmI[3].

Rationale

Future IT systems are getting more and more science embedded and user centered. Apart from specific businesses and scientific computations, systems have to provide more and more intelligence. An analysis of requirements revealed the lack of base research knowledge in information regulation principles of individual and social behaviour, data flow algorithms, hereditary IT design/solutions. The point is to develop and gain technologies, which, being in line with the determined ambitious perspectives of 6th FP will also become the key issue of future IT businesses. Experimentation with technologies under consideration is possible on base of knowledge available today: algorithms, adaptive and developing systems. This will produce approximate solutions. The effectiveness of such systems requires researches of specific type, which, in future, will make a background for the targeted next generation solutions – creative systems living and interacting both with other similar environments and with humans in information ambience. Being sophisticated, non-standard and presented within a theoretical framework, this R&D is within the scope of FET activities.

Impact

Industry: the move from local to distributed heterogeneous management is based on data flows, knowledge acquisition and knowledge accumulation. Such human like information systems will support the businesses providing services with effectiveness and with AmI properties.

Economy or society: distributed IT services and integrative indicators are the base components. For dynamic and flow data of huge sizes, systems will work effectively when several new paradigms are addressed to. For complex systems, the regular approaches will face shortage of resources: time, memory, algorithms and communication.

Communities addressed, other related initiatives

Research communities like philosophers, mathematicians, IT, psychologists, theoretical biologists, behavioural scientists, etc. may participate in HHS research. In described area initiatives exist both on European[4] and US[5]

---

[3] http://www.cordis.lu/ist/istag.htm

[4] Summary by John Casti and Ralph Dum, Brainstorming Meeting on Complex Systems, EC IST FET Unit and NOE EXYSTENCE, April 25-26, 2002, Brussels.

levels, although the existing initiatives are emphasizing elements different from those in HHS problem area. HHS is not only about classification (pattern recognition), machine learning (artificial intelligence), and complexities (discrete mathematics), it just addresses possibilities and approaches to systems, acting on informational level like human beings and communicating and developing like nature and society. Based on traditional intelligence studies it is searching constructions able to solve the integrated intelligent information management issues and even beyond, taking into account the exceptional power of computing machinery.

Contact Details

Prof. Levon Aslanyan from Information Society Technologies Center, Armenia; E-mail: lasl@sci.am

[5] Terascale and Petascale Computing: Digital Reality in the New Millennium, The Federal High-Performance Computing and Communications (HPCC) program (2003), The National Science Foundation, Arlington, Virginia 22230, USA.

FOURIER NEURAL NETWORKS: AN APPROACH WITH SINUSOIDAL ACTIVATION FUNCTIONS[6]

L.F de Mingo, L. Aslanyan, J. Castellanos, M. A. Díaz, V. Riazanov

Abstract: *This paper presents some ideas about a new neural network architecture that can be compared to a* Fourier *analysis when dealing periodic signals. Such architecture is based on sinusoidal activation functions with an* axo-axonic *architecture [1]. A biological axo-axonic connection between two neurons is defined as the weight in a connection in given by the output of another third neuron. This idea can be implemented in the so called* Enhanced Neural Networks *[2] in which two* Multilayer Perceptrons *are used; the first one will output the weights that the second* MLP *uses to computed the desired output. This kind of neural network has universal approximation properties [3] even with lineal activation functions.*

The only free parameters in the learning algorithm are the weights of one *MLP* since the weights of the other *MLP* are outputs computed by a neural network. This way the backpropagation algorithm must be modified in order to propagate the *Mean Squared Error* through both *MLPs*.

When all activation functions in an axo-axonic architecture are lineal ones ($f(x)=ax+b$) the output of the neural network is a polynomial expression in which the degree n of the polynomial depends on the number m of hidden layers [2] ($n=m+2$). This lineal architecture behaves like *Taylor* series approximation but with a global schema instead of the local approximation obtained by Taylor series. All boolean functions $f(x_1,\dots,x_n)$ can be interpolated with a axo-axonic architecture with lineal activation functions with n hidden layers, where n is the number of variables involve in the boolean functions. Any pattern set can be approximated with a polynomial expression, degree $n+2$, using an axo-axonic architecture with n hidden layers. The number of hidden neurons does not affects the polynomial degree but can be increased/decreased in order to obtained a lower *MSE*.

This lineal approach increases *MLP* capabilities but only polynomial approximations can be made. If non lineal activation functions are implemented in an axo-axonic network then different approximation schema can be obtained. That is, a net with sinusoidal functions outputs *Fourier* expressions, a net with *ridge* functions outputs ridge approximation, and so on. The main advantage of using a net is that a global approximation is achieved instead of a local approximation such as in the Fourier analysis.

In short, considering that only output neurons of the net computing weights have sinusoidal functions, then the output of this net is:

$$o_j = \sum_{i=0}^{N} w_{ij} o_i \text{ , where } f(x) = \sin(ax+b)$$

Taking into account that weights w_{Jl} of the other net are computing by previous output o_j, we can say that $w_{Jl}=o_j$, where $j=J*N+l$. Then, desired output follows equation:

$$o_k = \sum_{l=0}^{N_l} w_{kl} o_l \text{ , where } w_{kl}=w_{Jl}=o_j=sin(.)$$

[6] Supported by INTAS 2000-626, INTAS YSF 03-55-1969, INTAS INNO 182, and TIC2003-09319-c03-03.

Therefore, considering *p* hidden layers in the net, output can be expressed as:

$$o_k = \sum_{n=0}^{n=l+2} A \sin^n(.) + B$$

Considering previous equation, equations involved in *Fourier* analysis are similar to those obtained by axo-axonic networks.

Multilayer perceptrons are a subset of non lineal axo-axonic networks, since axo-axonic architectures with a given weights configuration behaves same way as a multilayer net. That is the reason why axo-axonic nets have more approximation capabilities than *MLP*. But the degree of output equation in a *MLP* can not be computed a priori such as in the axo-axonic architectures. Proposed architecture outputs a sinusoidal basis and a non lineal combination of it in order to obtained desired output and it can be compared to a *Fourier* analysis, moreover, Fourier coefficients can be used to initialize weights of neural networks in order to start the learning process with a low error.

Axo-axonic architectures can be used to forecast signals since they behave like *Fourier* analysis. This kind of connections can also be implemented on *Time-Delay* networks to improve results when dealing periodic signals. Some applications have been developed in order to forecast stock markets, weather and load demand.

Bibliography

[1] Mingo L.F., Diaz M.A., Gisbert F., Carrillo J.D.: **Enhanced Neural Networks: An Overview**. Knowledge-based Intelligent Information Engineering Systems and Allied Technologies. Frontiers in Artificial Intelligence and Applications. IOS Press Ohmsha. ISSN: 0922-6389. ISSN: 1535-6698. Vol.: 69. Part II. Pp.: 1170-1173. 2001.

[2] Luis F. Mingo, Miguel A. Diaz, Perez M., Palencia V.: **Enhanced Neural Networks as Taylor Series Approximation**. Knowledge-Based Intelligent Information Engineering Systems \& Allied Technologies. Frontiers in Artificial Intelligence and Applications. Vol. 82. IOS Press. ISSN: 0922-6389. Pp.: 532-533. 2002.

[3] de Mingo L.F., Gisbert P., Carrillo J., Hernandez C.: **Regression Hypersurfaces with a Modified Neural Network Architecture**. Advances in Soft Computing. ISSN 1615-3871. Pp: 241-249. 2003.

Author information

Mingo L.F: Dpto. Organización y Estructura de la Información - Escuela Universitaria de Informática
Universidad Politécnica de Madrid - Crta. de Valencia km. 7 - 28031 Madrid SPAIN; lfmingo@eui.upm.es
Aslanyan L.: Institute for Informatics and Automation Problems - NAS Armenia P. Sevak 1 Str.
Yerevan 14 Armenia; lasl@sci.am
Castellanos J.: Dpto. Inteligencia Artificial - Facultad de Informática - Universidad Politécnica de Madrid
28660 Boadilla del Monte - Madrid SPAIN; jcastellanos@fi.upm.es
Díaz M.A.: Dpto. Organización y Estructura de la Información - Escuela Universitaria de Informática
Universidad Politécnica de Madrid - Crta. de Valencia km. 7 - 28031 Madrid SPAIN; mdiaz@eui.upm.es
Riazanov V.: Computer Center of Russian Academy of Sciences, Vavilov, 40, Moscow; rvv@ccas.ru

MULTI-AGENT SYSTEMS IN THE HARVEST PROGNOSIS

Laura Ciocoiu, Cristian Paraschiv, Barbu Dragoş

Abstract: *The paper presents a case study of geo-monitoring a region consisting in the capturing and encoding of human expertise into a knowledge-based system. As soon as the maps have been processed, the data patterns are detected using knowledge-based agents for the harvest prognosis.*

Keywords: *data mining, topological maps, GIS, knowledge based agents, Model Based Reasoning*

Introduction

The process of geo-monitoring a region needs to use knowledge-based systems as a resource for aiding the specialists and the people in achieving their objectives. The model design process represents, in fact, the transferring of human experience in monitoring into an *interactive model.* Knowledge about process of the geo-monitoring of a region (e.g. map interpretation, statistics methods, strategies, etc) is represented by models that refer to observable features and significance.

So, the model designer has an inference network for the reasoning process; following the initial design of a model, the evaluation techniques are used to assess it in order to extent it; so, there are identified specific parts of the model that should be modified and the priorities of these modifications should be recorded. Besides the way of the geo-monitoring a region for a number of subjects, this process is in an on-going change (interference of Internet knowledge, including the artificial intelligence elements).

Area Monitoring

The aim of this work is to present a case study for capturing and encoding human expertise in geo-monitoring a region into a knowledge-based system. The main goals of weather and geo-monitoring of a region involving the environmental preservation are: modelling of liquid flows of rivers, harvest prognosis, riverbed degradation (utilization of hydropower potential) such some snapshots of our site "Geo-monitoring of a region" **http://inteligent-agents.ici.ro/gis/**

After the pre-processing operation of the topological and lithographic maps, knowledge about the process of the geo-monitoring of a region (e.g. statistics methods, strategies, etc) is represented by models that refer to observable features and significance.

The prognosis of harvest is realized on the base of geographical data such as using knowledge based modelling (hierarchical tree of classes and method definition, FBS decomposition, agent cooperation)

Harvest Prediction

This tool uses the knowledge engineering concepts such as *function-behaviour relationships, intelligent agents definitions* for creating a complex database for geo-monitoring a region.

Knowledge Representation is based on Model Based Reasoning (MBR) for an intuitive functional description, which is concerned with the design knowledge and geo-monitoring actions. The Functional-Behaviour-Structure (FBS) modelling consists of recursively decomposition of the functions into sub functions using a catalogue to look up the *most appropriate functional element* that means a component or a set of components that perform a function. Functional reasoning adds functional concepts into functional hierarchy and uses *a Casual Process description* with Casual Functional representation. The tool uses shared knowledge bases where the knowledge is represented as classes, objects and rules.

Function-Behaviour-State/Structure Modeller supports functional design (functional knowledge decomposition and function modelling) that means the defining of *function-behaviour* and *behaviour-state (structure)* relationships, causal decomposition (hierarchies constructed from the *function viewpoint*), construction of behaviour network, behaviour simulation, evaluation.

Knowledge Representation consists of constructing Function Knowledge Base and Behavior Knowledge Base based on model-based reasoning permit to generate a lot of alternatives, some of them "the best" for this process.

The process of geo-monitoring a region is simulated by constructing the Intelligent Agents referring to collaboration/ communication facilities. So, this model supports multiagents that generate partial results, data dependencies and conflicts among them.

Due to the complexity of application, in the 'Harvest prediction' process, along with the GIS software, a Knowledge-based System is used. First of all, there are defined superclasses **geographical features, geological structure, soil type, seed type, improvements, rainfall** with their attributes and instances. Also, some special data structures and classes had to be defined such as **Harvest, Basic Tilling, Auxiliary tilling, Soil Maintenance, Crop, Plant, Plant Control, Plant Maintenance**. External parameters (geographical position, climate), soil and seed, with their attributes (type, treatment, history, period) and their methods (work, evolution) are considered. The goal is quality and quantity of the crop. Each class represents an intelligent agent; the agents communicate each other in order to solve the conflicts.

For example, for a plain region with the quantity of the rainfall satisfactory and treatment, the harvest is expected to increase with 20%, according to the rules of the model.

The interpretation of the output data. We make the classes ***geographical features***, **rainfall** and **soil type**, which have the greatest influence on the harvest evolution. The most important class is ***geographical features,*** which determines the quantity of the rainfall and the type of the soil.

For the moment the model implementation is made for a few numbers of classes and rules. We intend to improve the algorithm to include additional classes and methods.

Conclusions and future trends

Collaborative efforts in artificial intelligence, statistics, and databases have been made in order to model the phenomena and to be a support of decision-making.

The digital geographic databases are simple representation of geographic knowledge at the level of basic geometric, topological and measurement constraints. Knowledge-based GIS attempts to build higher-level geographic knowledge into digital geographic databases for analysing complex phenomena. Geographic knowledge discovery is a source for knowledge-based GIS and intelligent analysis.

For the next work we intend to develop our application by map interpretation and information extraction, environmental features defining (e.g. inductive learning techniques and neural networks to classify and map soil types, mapping vegetation types), extracting patterns, creating the geographical database.

Acknowledgement

This Technical Report was made by a team from National Institute for Research and Development in Informatics, on September, 2003, under INTAS Project no. 397: "**Data Mining Technologies and Image Processing: Theory and Applications**" coordinated by Lappeenranta University of Technology, Finland.

We want to thank the team leader dr.mat. Angela Ionita, Deputy Director of RACAI, for her work and assistance during this period.

References

BHANDARI, E. COLET, E., PARKER, J., PINES, Z, PRATAP, R., PRATAP, R., RAMANUJAM, K., 1997, **Advanced scout: data mining and knowledge discovery in NBA data**. Data Mining and Knowledge Discovery, 1;

BRACHMAN, R.J., ANAND, T., 1996, **The Process of Knowledge Discovery in Databases**, in Advances in Knowledge Discovery and data Mining, U.M. Fayyad et al. (eds.) AAAI Press/ The MIT Press

BUTTENFIELD, B., GAHEGAN, M., MILLER, H., YUAN, M., **Geospatial Data Mining and Knowledge Discovery**;

FAYYAD, U. 1997. Editorial. **Data Mining and Knowledge Discovery**.

FOTHERINGHAM, A. S., CHARLTON, M. BRUNSDON, C., 1997, **Two techniques for exploring non-stationarity in geographical data**. Geographical Systems;.

GAHEGAN, M., 2000, **On the application of inductive machine learning tools to geographical analysis**, Geographical Analysis;

GAHEGAN, M. O'BRIEN, D. (1997). **A Strategy and Architecture for the Visualisation of Complex Geographical Datasets**. International Journal of Pattern Recognition and Artificial Intelligence, 11(2);

GAHEGAN, M., TAKATSUKA, M., WHEELER, M., HARDISTY, F., 2000, **GeoVISTA *Studio*: A Geocomputational Workbench**. Proc. 4th Annual Conference on GeoComputation, UK, August 2000. URL: http://www.ashville.demon.co.uk/gc2000/.

Authors' information

Laura Ciocoiu – senior research. Main interests are in intelligent agents, multi-agent systems, multimedia systems and security tehnology

Barbu Dragoş – research. Main interests are in intelligent agents, multi-agent systems, neural networks

Cristian Paraschiv – programmer. Main interests are in security tehnology and dynamic Web portals.

National Institute for Research and Development in Informatics

8-10 Averescu Avenue, 011455 Bucharest 1 ROMANIA

email: ciocoiu@ici.ro   http://intelligent-agents.ici.ro

PAPERS in ENGLISH

EXPERIENCE WITH WEB-BASED EDUCATIONAL PLATFORM – POSSIBILITIES AND LIMITATIONS

Petya Assenova, Georgy Todorov

1. Abstract

The main principles of building the NBU system are presented in the paper. The authors describe the functions, structure of the educational materials, possibilities for application and future development. An example of educational materials via VEDA is shown on the bachelor course "Programming and Data Structures". Some problems are discussed and comparative statistical data are presented.

2. About NBU

NEW BULGARIAN UNIVERSITY (NBU) is established in 1991 as a private university. Its main priorities are:

- development of the interdisciplinary programs,
- ensuring mobility to both students and instructors,
- development of new educational programs,
- exploration of new professional opportunities.

The programs for bachelor's, master's and doctor's degrees are based on a credit system. There are several forms of education:

- full-time education,
- distant education,
- continuing education.

Mainly NBU offers education with a humanitarian profile. Our program "Informatics and Telecommunications" is one of the few science programs.

3. History of the web-based education in NBU – VEDA

We can divide our university's experience in the field of web-based education in two periods.

The First period began with a project called "Virtual Education for Master Programs" financed by Open Society Foundation in 2002 and launched by the Computer Science Department. Both instructors and students were involved in the project. The instructors played mainly supervisor and counseling role, and the students did the design, coding and testing. One year later the system for web-based education was completed and it included the following components:

- Courses and lectures,
- Assignments,
- Self-test system,
- Glossary

The potential users of that system were:

- Full-time students,
- Students from the distant-learning department
- Part-time students
- Instructors
- Administrator (this system needed an administrator who would manually create accounts, course list, specify instructors and students for each course, etc)

The system was used for a relatively short period of time, because the NBU Council Board asked to integrate that system with the main information system of the University. This lead to the start of the **second period** of our project.

4. What is VEDA

As a result of the second stage a system called VEDA (originally from Virtual Education Access) was developed. This name has a symbolic meaning – the old Slavonic word "veda" stands for "knowledge".

Now VEDA is accessible to each student and instructor of our University via the NBU web-page. It is used on a regular basis by instructors and students.

VEDA can be characterized with the following main features:

- **Integration with the university information system**. It means that the students' and instructors' data is obtained automatically from the information system (this includes list of courses, course details, students list, etc – the information is divided by terms).
- **Main features of the educational materials and functions:** course description (short professional data for the lectures, objectives and competencies, syllabus, references), lectures, materials attached to the lecture, assignments, self-tests, forum, statistics.

All teaching materials can be uploaded in different formats: text, pdf, powerpoint, graphic files (all other files can be uploaded as attachments to a lecture). The materials can be uploaded by a variety of methods – typing directly from the keyboard, using the copy-paste option and as files, using the browse option. The pasted text preserves its layout, as it is originally in a Word or some other file or some other text-processing program. The course materials can be downloaded or printed, edited or removed.

The structure of all uploaded data which is uploaded, follows the requirements of the university, thus enabling VEDA not only to fetch data from the main university database, but also to provide it with data.

The assignments module enables instructors to set start and deadline dates for submitting assignments, as well as to provide additional guidelines regarding the assignment. Students submit their assignments and can see their grades on-line.

At this time, the self-test system, is only capable of creating tests with questions which have one or multiple correct answers (true or false). Each correct answer gives a number of points, and the sum of the points determines the final grade.

Using the statistics module, each instructor can check how much time their students spend on reading lectures or on taking self-tests.

For each course, there is a discussion board. Every student can participate in the discussion boards of the courses they are enrolled in. The moderator of the discussion is the instructor.

- **Tools for developing course materials**. Instructors can develop their courses in any format they like. Natively, the system supports plain-text, HTML, DOC and RTF, PDF and PowerPoint. All other formats, such as audio or video, Flash, etc, can be uploaded as attachments, for the students to download. Creating and uploading a lecture is easy, and only requires the completion of a five step wizard interface. At any step, there is a help text available and also a real person contact information.
- **Friendly interface and help for users**. When designing the user interface we had to take into consideration several factors. Most importantly, the system had to be user-friendly and intuitive. Without this, instead of serving its users, it would frustrate them, and very soon would prove to be unusable. As mentioned in the introduction, NBU mainly provides humanitarian education – hence we expected that most of the users would not be technically savvy. Another important issue was that most of the users would be accessing the system remotely, and most of them would use dialup connections. After reviewing several proposals we selected a design which uses few clip-art pictures. We also implemented a module, which sends compressed content to the users whose browsers can automatically decompress it, thus further improving the speed. We have implemented a short-cut menu, which allows users to quickly jump through the most recently visited pages. And for those who need help, every page has a 'Help' link, which contains a contact form and help text, relevant to the page the user has opened.

5. We plan developments in the following fields:

- **Connecting VEDA to the electronic resources of the university library;**

Subject of another concurrent project is the digitalisation of rare or single-copy books and magazines in possession of the University Library. The idea we are currently working on is the integration of these digital resources with VEDA. This way, instructors would be able not only to provide a plain list of text books for their course, but also to select materials directly from a web interface, thus providing their students a one-click access to reading materials. The launching of this module is scheduled for July.

- **More possibilities to format lecture materials**

This would include more advanced layout formatting tools available on-line, and also improvements of the security and modifications according to the input we are receiving from the users.

- **Improving the self-test module**

The short term plans include the adding new types of test questions and answers -- e.g. Questions and answers which include clipart, 'fill in the blank' type answer, etc. Another activity relates to developing the glossary from first version and to includes it in the new system.

In the long-term we plan to make our system compatible with other systems and standards for web-based education.

6. Some examples with VEDA

After login in the system instructors can chose the term and school year. The list of his/her courses is opened. The instructor may open the course to develop new materials, edit existing materials, read, print or download them (Fig. 1.).

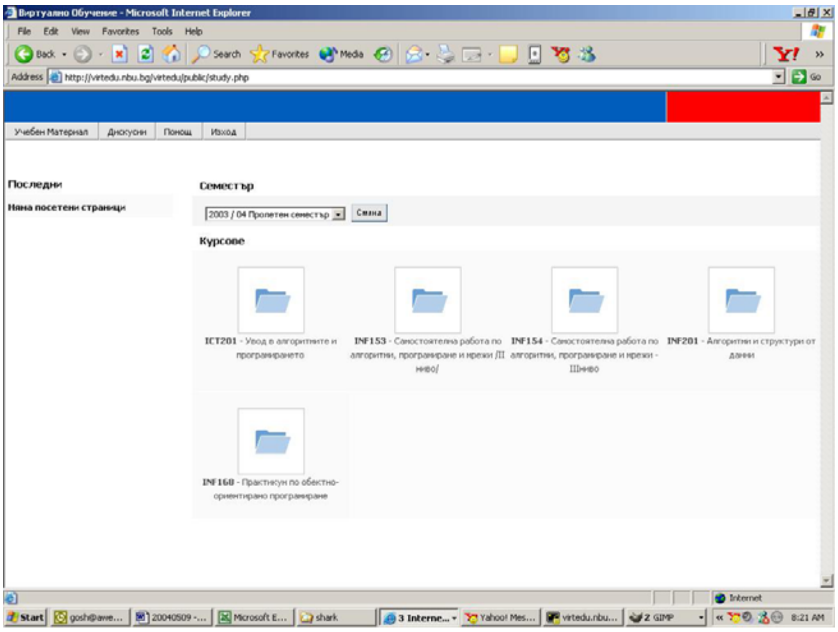


Fig.1. The personal instructor page.

The fig. 2. presents the screen where instructor can develop and upload one of the lecture element – objectives. If he/she needs help how to do this, the help is available on the screen – just read it. After adding this element, the instructor go to next element (button "ahead").

Fig. 3 shows the view to a lecture from the course "Algorithms and Data Structures" in pdf format. A short-cut menu is shown for quick jumps through the recently visited pages.

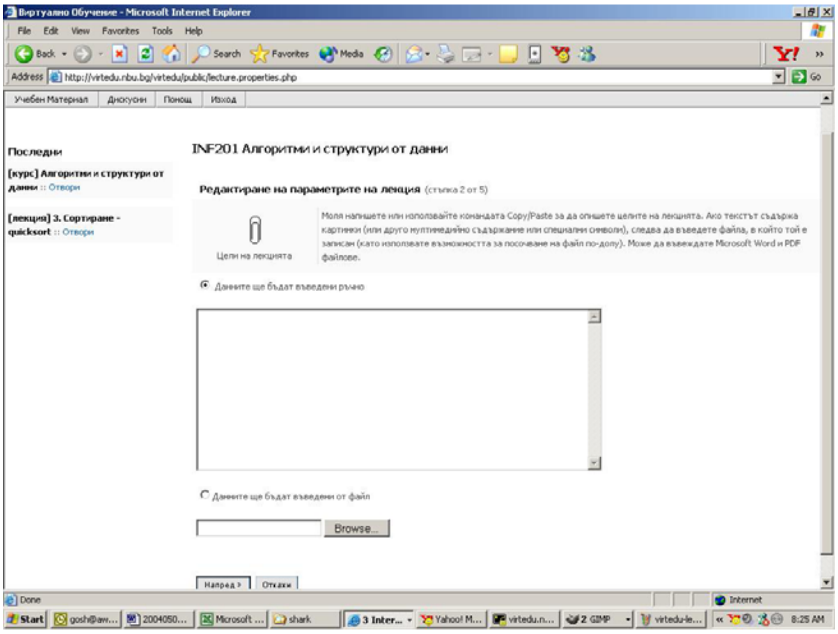


Fig. 2. Window for developing objectives of the lecture.

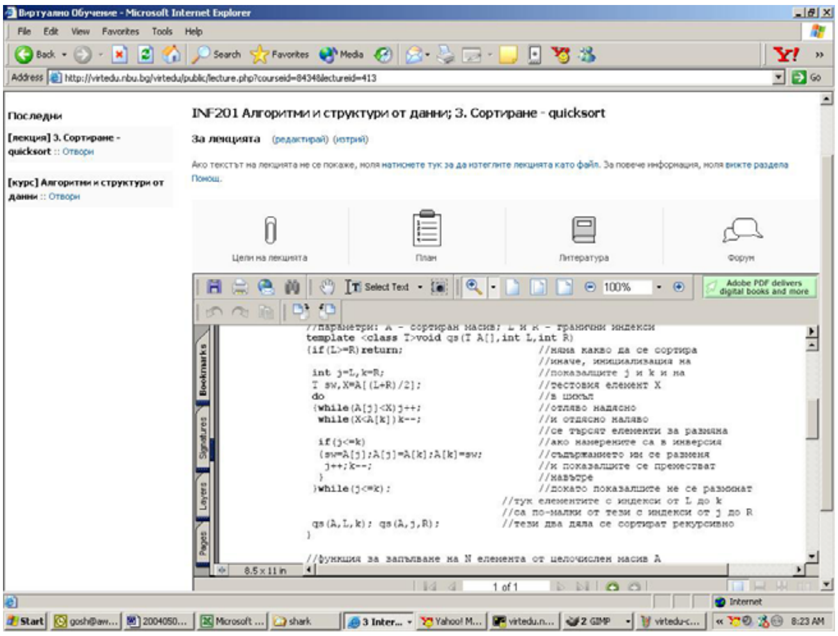


Fig. 3. The window for the lecture.

The interface in the student's pages looks the same way but students can not the writes to change the materials.

7. Problems

During the short period of time the system has been running we have faced a number of problems. By far, the main trouble comes from the lack of motivation of the instructors to spend additional time to prepare and upload their digital materials. There are several reasons for this including the instructors' lack of computer skills and the additional time they would need to spend in order to prepare their materials. Fortunately the NBU council board passed a regulation according to which, instructors who make available their lectures online using VEDA will receive bonuses. This proposal was accepted a week ago, and we hope this fact will significantly increase instructors' interest in VEDA.

Another important problem is protection of the copyright rights of the documents which are published on-line. Consider the scenario where a student from NBU downloads all materials for a course he has access to and uploads it to the internet for everyone to download, or worse, starts selling it to students from other universities. In Bulgaria, where copying of lecture materials (using a conventional copier machine) is quite common, this problem presents one of the biggest obstacles we need to overcome.

And finally, a purely technical problem. Rewiring the university network and changing configurations led to accessibility problems.

8. Some statistics

Fig. 4. shows the number of instructors who have uploaded any materials in VEDA. The measures are made 3 months after starting VEDA. The number of instructors increases before the session. This data presents 25 percents of all full-staff professors and 2 % of all professors delivering lectures in NBU.

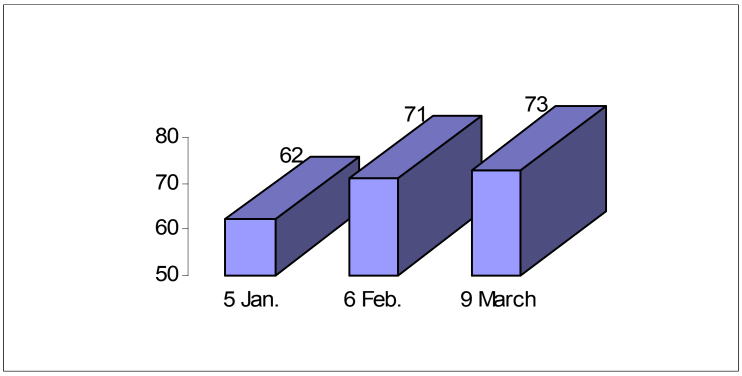


Fig. 4. The instructors' activity

The students' demand is much more then lecturers supply. There are 2.000 students' queries to the server in January, a little more - in February and more then 4.200 – in March. The number of all students in NBU is 8.000. The data shows that the students need educational materials via Internet but the lecturers' activity is not satisfied.

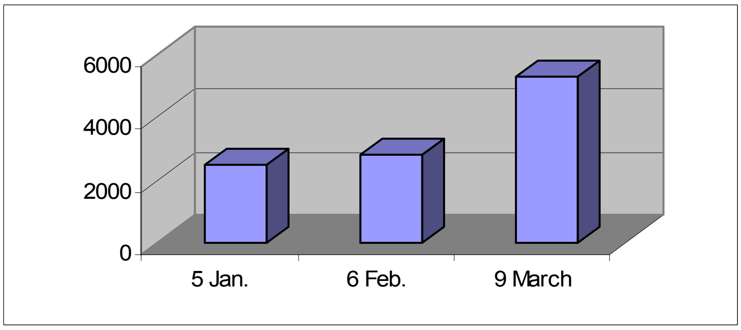


Fig. 5. The students' activity

Fig. 6 presents the number of lectures from different subjects uploaded in VEDA.

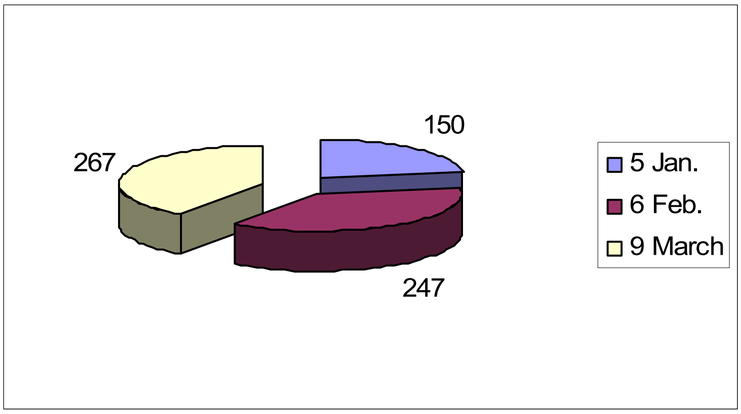


Fig. 6. Lectures in VEDA.

9. Conclusion

In conclusion I can say that we are confident that VEDA will be of great benefit to instructors, who will have an easy and convenient way to keep in touch with their students, monitor their progress and distribute assignments.

Students, on the other hand, will have a centralized real-time access to the lecture materials, an e-library and most importantly, an opportunity to contact and discuss questions directly with their instructors, regardless of where they are.

This project will significantly improve the quality of the education process at NBU, and will enable our university to target a much wider array of prospective students, e.g. People who want to obtain master's degree, but are currently unable to do so, because their everyday schedule prevents them from attending lectures. Up until now, such individuals had limited choice, and had to enrol in foreign universities, specialized in distant learning.

Starting next semester, our department will also offer an English language program. Combined with VEDA, this new program will also target foreign students from near-by countries. Thus they would only need to come to campus to take exams.

Having said the above, we are looking forward to the official roll-out, which will take place in September, at the start of the next fall semester.

10. Bibliography

1. Virtual University. http://vu.org/
2. VEDA. http://wwww.nbu.bg/

Authors' information

Petya Assenova: Assoc. Prof. Ph.D., Head of Computer Science Department, New Bulgarian University, 21 Montevideo Str., Sofia – 1618, Bulgaria, e-mail: pasenova@nbu.bg

Georgi Todorov: student, Department of Computer Science, New Bulgarian University, 21 Montevideo Str., Sofia – 1618, Bulgaria, e-mail: gtodorov@nbu.bg

PROJECT SOLUTIONS REUSE

M. Bondarenko, A. Kuzemin, M. Sorochan, I. Yanchevskiy

Extended Abstract:

Current information technologies development makes it possible to solve a wide range of tasks in different types of organizations. Modern software development process is becoming more and more complex. Before developing software product designers have to define the project goal. During that process they may find the problem of misunderstanding between designers and future customers of software being developed. Modern CASE software at most are only visualization tools and have no support of the project solutions reuse. Mostly reuse is known in the component development. But research of development process shown that any project artifact may be reused, e.g. diagrams, quizzes, rules, instructions and so on.

The main problem in the solutions reuse is the process of searching for the most suitable one. Such task definition comes from the system environment description and defining its goals.

Solution is in the development of a methodology for the design solutions reuse by means of a language of situation representation. The main idea consists in connection of a system objective with situation built from entities which describe the condition of the system at the time of objective establishing. Every situation is associated with one or several design solutions, which can be used in the process of the development. The language of situation representation has been built. This makes it possible to describe a problem situation with a natural language. To compare situations between each others similarity measure was developed based on similarity coefficients with an absent part value added.

The goal of software being developed is to solve some goal of the system product for which it is made. The goal is connected with the system and its environment conditions. So one may say that he has such a chain system condition – system goal – solution. Having some system condition we can get some solution for that condition. Such conditions are described by situations. To find definite solution n the developed methodology one has to build a situation for which he is looking for solutions for.

Situation consists of microsituations each of them is triple of three entities. They are central entity, context and secondary entity. The central one is the object action or its description. Context defines the state of the central entity or relation with a secondary one. The secondary entity is the object central entity act with.

To build a situation one can define the entities it first consists of. All entities are divided into categories. Each category is the set of entities grouped by some feature and similarity. E.g. organization structures, act types, financials, actor relations and so on. Entities in category are organized by generalization. At the top of the category the first entity is. Other entities are generalized from it. Each child entity has all the features and properties of the parent one.

The search process is built on similarity measure between etalon and problem situations which is based on Tversky similarity model with an absent part value. Similarity is counted for situation, microsituations and entities they are built on. For situations the measure is based on comparing microsituations. Their measure is based on entities comparison. Entity being compared is based on categories structure and generalization associations between them. Special algorithms are developed to find the described measures.

Authors' Information

Bondarenko M.F.: Prof., Rector of Kharkov National University of Radio Electronics

Kuzemin A.Ya.: Prof. of Information Department, Kharkov National University of Radio Electronics

Sorochan M.V.: post-graduate student of Kharkov National University of Radio Electronics, msorc@ua.fm

Yanchevskiy I.V.: General Director of the Joint Russian-Kirghiz Venture "Computing Technique and Automation Means"

REPRESENTING "RECURSIVE" DEFAULT LOGIC IN MODAL LOGIC

Frank M. Brown

Abstract: *The "recursive" definition of Default Logic is shown to be representable in a monotonic Modal Quantificational Logic whose modal laws are stronger than S5. Specifically, it is proven that a set of sentences of First Order Logic is a fixed-point of the "recursive" fixed-point equation of Default Logic with an initial set of axioms and defaults if and only if the meaning of the fixed-point is logically equivalent to a particular modal functor of the meanings of that initial set of sentences and of the sentences in those defaults. This is important because the modal representation allows the use of powerful automatic deduction systems for Modal Logic and because unlike the original "recursive" definition of Default Logic, it is easily generalized to the case where quantified variables may be shared across the scope of the components of the defaults.*

Keywords: *Recursive Definition of Default Logic, Modal Logic, Nonmonotonic Logic.*

1. Introduction

One of the most well known nonmonotonic logics [Antoniou 1997] which deals with entailment conditions in addition to possibility conditions in it's defaults is the so-called Default Logic [Reiter 1980]. The basic idea of Default Logic is that there is a set of axioms Γ and some non-logical default "inference rules" of the form:

$$\frac{\alpha : \beta_1 ... \beta_m}{\chi}$$

which is intended to suggest that χ may be inferred from α whenever each $\beta_1,...,\beta_m$ is consistent with everything that is inferable. Such "inference rules" are not recursive and are circular in that the determination as to whether χ is derivable depends on whether β_i is consistent which in turn depends on what was derivable from this and other defaults. Thus, tentatively applying such inference rules by checking the consistency of $\beta_1,...,\beta_m$ with only the current set of inferences produces a χ result which may later have to be retracted. For this reason inferences in a nonmonotonic logic such as Default Logic are essentially carried out not in the original nonmonotonic logic, but rather in some (monotonic) metatheory in which that nonmonotonic logic is monotonically defined. [Reiter 1980] explicated the above intuition by defining Default Logic "recursively" in terms of the set theoretic proof theory metalanguage of First Order Logic (i.e. FOL) with (more or less) the following fixed point expression[7]:

$$'\kappa=(\text{dr}\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)$$

where dr is defined as:

$(\text{dr}\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i) =_{df} \cup_{t=1,\omega}(\text{r}\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)$

$(\text{r}\ 0\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i) =_{df} (\text{fol}\ '\Gamma)$

$(\text{r}\ t{+}1\ 'k\ \ '\alpha_i{:}'\beta_{ij}/'\chi_i) =_{df} (\text{fol}((\text{r}\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(\text{r}\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin '\kappa)\}))$

where $'\alpha_i$, $'\beta_{ij}$, and $'\chi_i$ are the closed sentences of FOL occuring in the ith "inference rule" and $'\Gamma$ is a set of closed sentences of FOL. A closed sentence is a sentence without any free variables. fol is a function which produces the set of theorems derivable in FOL from the set of sentences to which it is applied. The quotations

---

[7] [Reiter 1980] actually used a recursive definition whereby the r sets do not necessarily contain all their FOL consequences:

$(\text{dr}\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i) =_{df} \cup_{t=1,\omega}(\text{r}\ \ t\ '\kappa)$

$(\text{r}\ 0\ \ '\kappa) =_{df}\ '\Gamma$

$(\text{r}\ t{+}1\ '\kappa) =_{df} (\text{fol}(\text{r}\ '\kappa\ t))\cup\{'\chi_i{:}('\alpha_i\varepsilon(\text{r}\ '\kappa\ i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin '\kappa)\}$

If this definition were used then all the theorems in this paper should have $(\text{r}\ \ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)$ replaced by $(\text{fol}(\text{r}\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ and $(\text{dr}\ \ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)$ replaced by $(\text{fol}(\text{dr}\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$.

appended to the front of these Greek letters indicate references in the metalanguage to the sentences of the FOL object language. Interpreted doxastically this fixed point equation states:

The set of closed sentences which are believed is equal to
the union of all sets of closed sentences which are believed at any time.
That which is believed at time 0 is the set of closed sentences derived by the laws of FOL from 'Γ.
That which is believed at time t+1 is the set of closed sentences derived by the laws of FOL
from the union of both
the set of beliefs at time t
and the set of all 'χ_i for each i such that
the closed sentence 'α_i is believed at time t and for each j, the closed sentence 'β_{ij} is believable.

The purpose of this paper is to show that all this metatheoretic machinery including the formalized syntax of FOL, the proof theory of FOL, the axioms of set theory, and the set theoretic fixedpoint equation is not needed and that the essence of the "recursive" definition of Default Logic in representable as a necessary equivalence in a simple (monotonic) Modal Quantificational Logic. Interpreted as a doxastic logic this necessary equivalence states:

That which is believed is logically equivalent to what is believed at any time.
That which is believed at time 0 is Γ.
That which is believed at time t+1 is
that which is believed at time t and for each i, if α_i is believed at time t and for each j, β_{ij} is believable then χ_i.

thereby eliminating all mention of any metatheoretic machinery.

The remainder of this paper proves that this modal representation is equivalent to the "recursive" definition of Default Logic. Section 2 describes a formalized syntax for a FOL object language. Section 3 describes the part of the proof theory of FOL needed herein (i.e. theorems FOL1-FOL10). Section 4 describes the Intensional Semantics of FOL which includes laws for meaning of FOL sentences: M0-M7, theorems giving the meaning of sets of FOL sentences: MS1, MS2, MS3, and laws specifying the relationship of meaning and modality to the proof theory of FOL (i.e. the laws R0, A1, A2 and A3 and the theorems C1, C2, C3, and C4). The modal version of the "Recursive" definition of Default Logic, called DR, is defined in section 5 and explicated with theorems MD1-MD8 and SS1-SS2. In section 6, this modal version is shown by theorems R1, DR1 and DR2 to be equivalent to the set theoretic fixed point equation for Default Logic. Figure 1 outlines the relationship of all these theorems in producing the final theorems DR2, FOL10, and MD8. .

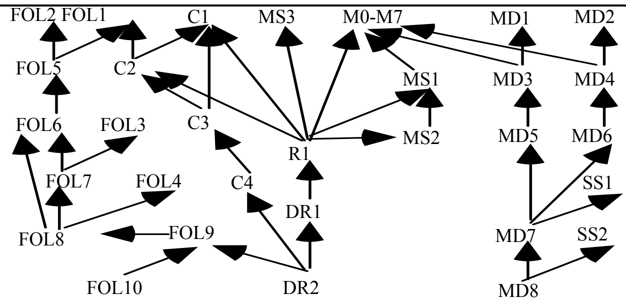


Figure 1: Dependencies among the Theorems

2. Formal Syntax of First Order Logic

We use a First Order Logic (i.e. FOL) defined as the six tuple: ($\rightarrow$, #f, $\forall$, *vars*, *predicates*, *functions*) where $\rightarrow$, #f, and $\forall$ are logical symbols, *vars* is a set of variable symbols, *predicates* is a set of predicate symbols each of which has an implicit arity specifying the number of associated terms, and *functions* is a set of function symbols each of which has an implicit arity specifying the number of associated terms. The sets of logical symbols, variables, predicate symbols, and function symbols are pairwise disjoint. Lower case Roman letters possibly indexed with digits are used as variables. Greek letters possibly indexed with digits are used as syntactic metavariables. γ, $\gamma_1,...\gamma_n$, range over the variables, ξ, $\xi_1...\xi_n$ range over sequences of variables of an appropriate arity, $\pi,\pi_1...\pi_n$ range over the predicate symbols, ϕ, $\phi_1...\phi_n$ range over function symbols, δ, $\delta_1...\delta_n$, σ range over terms, and α, $\alpha_1...\alpha_n$, β, $\beta_1...\beta_n,\chi$, $\chi_1...\chi_n$, $\Gamma_1,...\Gamma_n,\varphi$ range over sentences. The terms are of the forms γ and (ϕ δ1...δn), and the sentences are of the forms ($\alpha\rightarrow\beta$), #f, ($\forall\gamma$ α), and (π $\delta_1...\delta_n$). A nullary predicate π or function ϕ is written as a sentence or a term without parentheses. $\varphi\{\pi/\lambda\xi\alpha\}$ represents the replacement of all occurrences of π in φ by $\lambda\xi\alpha$ followed by lambda conversion. The primitive symbols are shown in Figure 2 with their intuitive interpretations.

Symbol	Meaning
$\alpha \rightarrow \beta$	if α then β.
#f	falsity
$\forall\gamma\ \alpha$	for all γ, α.

Figure 2: Primitive Symbols of First Order Logic

The defined symbols are listed in Figure 3 with their definitions and intuitive interpretations.

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$\neg\alpha$	$\alpha \rightarrow$ #f	not α		$\alpha\wedge\beta$	$\neg(\alpha \rightarrow \neg\beta)$	α and β
#t	$\neg$ #f	truth		$\alpha\leftrightarrow\beta$	$(\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$	α if and only if β
$\alpha\vee\beta$	$(\neg\alpha)\rightarrow\beta$	α or β		$\exists\gamma\ \alpha$	$\neg\forall\gamma\neg\alpha$	for some γ , α

Figure 3: Defined Symbols of First Order Logic

The FOL object language expressions are referred in the metalanguage (which also includes a FOL syntax) by inserting a quote sign in front of the object language entity thereby making a structural descriptive name of that entity. Generally, a set of sentences is represented as: {'Γ_i} which is defined as: {'Γ_i: #t} which in turn is defined as: {s: $\exists$i(s='Γ_i)} where i ranges over some range of numbers (which may be finite or infinite). With a slight abuse of notation we also write 'κ, 'Γ to refer to such sets.

3. Proof Theory of First Order Logic

FOL is axiomatized with a recursively enumerable set of theorems as its axioms are recursively enumerable and its inference rules are recursive. The axioms and inference rules of FOL [Mendelson 1964] are given in Figure 4.

MA1: $\alpha \rightarrow (\beta\rightarrow \alpha)$

MA2: $(\alpha\rightarrow (\beta\rightarrow \rho)) \rightarrow ((\alpha\rightarrow \beta)\rightarrow (\alpha\rightarrow \rho))$

MA3: $((\neg \alpha)\rightarrow (\neg \beta))\rightarrow (((\neg \alpha)\rightarrow \beta)\rightarrow\alpha)$

MA4: $(\forall\gamma\ \alpha)\rightarrow \beta$ where β is the result of substituting an expression (which is free for the free positions of γ in α) for all the free occurrences of γ in α.

MA5: $((\forall\gamma(\alpha \rightarrow \beta)) \rightarrow (\alpha\rightarrow(\forall\gamma\ \beta)))$ where γ does not occur in α.

MR1: from α and $(\alpha\rightarrow \beta)$ infer β

MR2: from α infer $(\forall\gamma\ \alpha)$

Figure 4: Inferences Rules and Axioms of FOL

In order to talk about sets of sentences we include in the metatheory set theory symbolism as developed along the lines of [Quine 1969]. This set theory includes the symbols ε, $\not\in$, $\supseteq$, $=$, $\cup$ as is defined therein. The derivation operation (i.e. fol) of any First Order Logic obeys the Inclusion (i.e. FOL1), Idempotence (i.e. FOL2), Monotonic (i.e. FOL3) and the Union (i.e. FOL4) properties:

FOL1: (fol 'Γ)$\supseteq$'Γ Inclusion

FOL2: (fol 'κ)$\supseteq$(fol(fol 'κ)) Idempotence

FOL3: ('$\kappa\supseteq$'Γ) $\rightarrow$ ((fol 'κ)$\supseteq$ (fol 'Γ)) Monotonicity

FOL4: For any set ψ, if $\forall$t((ψ t)=(fol(ψ t))) and $\forall$t((ψ t+1)$\supseteq$(ψ t)) then ($\cup_{t=0,\omega}$(ψ t))=(fol($\cup_{t=0,\omega}$(ψ t)) Union

From these four properties we prove the following theorems of the proof theory of First Order Logic:

FOL5: $(fol\ '\kappa)=(fol(fol\ '\kappa))$ proof: By FOL1 and FOL2.

FOL6: $(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)=(fol(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ proof: By induction on t it suffices to prove:

(1) $(r\ 0\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)=(fol(r\ 0\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ Unfolding r twice gives: $(fol\ '\Gamma)=(fol(fol\ '\Gamma))$ which is FOL5.

(2) $(r\ t{+}1\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)=(fol(r\ t{+}1\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$

Unfolding r twice gives: $(fol((r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\}))$

$=(fol((fol((r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\}))$

which likewise is an instance of FOL5. QED.

FOL7: $((r\ t{+}1\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\supseteq(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)))$

proof: By FOL6 this is equivalent to: $((r\ t{+}1\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\supseteq(fol(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))))$. Unfolding r of t+1 gives:

$(fol((r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\}))\supseteq(fol(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$

By FOL3 it suffices to prove: $((r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\}))\supseteq(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)$ which holds in set theory. QED.

FOL8: $(\cup_{t=0,\omega}(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))=(fol(\cup_{t=0,\omega}(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)))$

proof: $\forall t((r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))=(fol(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)))$ holds by FOL6. $\forall t((r\ t{+}1\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\supseteq(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ holds by FOL7. Instantiating the hypotheses in FOL4 to these theorems proves this theorem. QED.

FOL9: $(dr\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)=(fol(dr\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ proof: Unfolding dr twice gives: $\cup_{t=1,\omega}(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i)=(fol(\cup_{t=1,\omega}(r\ t\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))$ which holds by FOL8. QED.

FOL10: $('k=(dr\ '\kappa\ '\Gamma\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\rightarrow('\kappa=(fol\ '\kappa))$ proof: From the hypothesis and FOL9 $'k=(fol(dr\ '\kappa))$ is derived. Using the hypothesis to replace $(dr\ '\kappa)$ by $'\kappa$ in this result gives: $('\kappa=(fol\ '\kappa))$ QED.

4. Intensional Semantics of FOL

The meaning (i.e. mg) [Brown 1978, Boyer&Moore 1981] or rather disquotation of a sentence of FOL is defined in Figure 5 below[8]. mg is defined in terms of mgs which maps each FOL object language sentence and an association list into a meaning. mgn maps each FOL object language term and an association list into a meaning. An association list is a list of pairs consisting of an object language variable and the meaning to which it is bound.

M0: $(mg\ '\alpha) =df\ (mgs\ '(\forall\gamma1...\gamma n\ \alpha)'())$ where $'\gamma1...'\gamma n$ are all the free variables in $'\alpha$

M1: $(mgs\ '(\alpha\rightarrow\beta)a)\leftrightarrow((mgs\ '\alpha\ a)\rightarrow(mgs\ '\beta\ a))$

M2: $(mgs\ '\#f\ a)\leftrightarrow\#f$

M3: $(mgs\ '(\forall\ \gamma\ \alpha)a)\leftrightarrow\forall x(mgs\ '\alpha(cons(cons\ '\gamma\ x)a))$

M4: $(mgs\ '(\pi\ \delta_1...\delta_n)a)\leftrightarrow(\pi(mgn\ '\delta_1\ a)...(mgn\ '\delta_n\ a))$ for each predicate symbol $'\pi$.

M5: $(mgn\ '(\phi\ \ \delta_1...\delta_n)a) = (\phi(mgn\ '\delta_1\ a)...(mgn\ '\delta_n\ a))$ for each function symbol $'\phi$.

M6: $(mgn\ '\gamma\ a) = (cdr(assoc\ '\gamma\ a))$

M7: (assoc v L)= (if(eq? v(car(car L)))(car L)(assoc v(cdr L))) where: cons, car, cdr, eq?, and if are as in Scheme.

Figure 5: The Meaning of FOL Sentences

The meaning of a set of sentences is defined in terms of the meanings of the sentences in the set as:

$(ms\ '\kappa) =df\ \forall s((s\varepsilon'\kappa)\rightarrow(mg\ s))$.

MS1: $(ms\{'\alpha{:}\ \Gamma\})\leftrightarrow\forall\xi((\Gamma\{s/'\alpha\})\rightarrow\alpha)$ where ξ is the sequence of all the free variables in $'\alpha$ and where Γ is any sentence of the intensional semantics. proof: $(ms\{'\alpha{:}\Gamma\})$ Unfolding ms and the set pattern abstraction symbol

[8] The laws M0-M7 are analogous to Tarski's definition of truth except that finite association lists are used to bind variables to values rather than infinite sequences. M4 is different since mg is interpreted as being meaning rather than truth.

gives: $\forall s((s\varepsilon\{s: \exists\xi((s='\alpha)\wedge\Gamma)\})\rightarrow(mg\ s))$ where ξ is a sequence of the free variables in 'a. This is equivalent to: $\forall s((\exists\xi((s='\alpha)\wedge\Gamma)))\rightarrow(mg\ s))$ which is: $\forall s\forall\xi\ (((s='\alpha)\wedge\Gamma)\rightarrow(mg\ s))$ which is: $\forall\xi(\Gamma\{s/'\alpha\}\rightarrow(mg\ '\alpha))$. Unfolding mg using M0-M7 then gives: $\forall\xi((\Gamma\{s/'\alpha\})\rightarrow\alpha)$ QED

The meaning of a set is the meaning of all the sentences in the set (i.e. MS2):

MS2: $(ms\{'\Gamma_i\}) \leftrightarrow \forall i\forall\xi_i\Gamma_i$ proof: $(ms\{'\Gamma_i\})$ Unfolding the set notation gives: $(ms\{'\Gamma_i: \#t\})$. By MS1 this is equivalent to: $\forall_i\forall\xi_i((\#t\{s/'\alpha\})\rightarrow\Gamma_i)$ which is equivalent to: $\forall i\forall\xi_i\Gamma_i$ QED.

The meaning of the union of two sets of FOL sentences is the conjunction of their meanings (i.e. MS3):

MS3: $(ms('\kappa\cup'\Gamma)) \leftrightarrow ((ms\ '\kappa)\wedge(ms\ '\Gamma))$ proof: Unfolding ms and union in: $(ms('\kappa\cup'\Gamma))$ gives:

$\forall s((s\varepsilon\{s: (s\varepsilon'\kappa)\vee(s\varepsilon'\Gamma)\})\rightarrow(mg\ s))$ or rather: $\forall s(((s\varepsilon'\kappa)\vee(s\varepsilon'\Gamma))\rightarrow(mg\ s))$ which is logically equivalent to: $(\forall\alpha((s\varepsilon'\kappa)\rightarrow(mg\ s)))\wedge(\forall s((s\varepsilon'\Gamma)\rightarrow(mg\ s)))$. Folding ms twice then gives:$((ms\ '\kappa)\wedge(ms\ '\Gamma))$ QED.

The meaning operation may be used to develop an Intensional Semantics for a FOL object language by axiomatizing the modal concept of necessity so that it satisfies the theorem:

C1: $('\alpha\varepsilon(fol\ '\kappa)) \leftrightarrow ([]\ ((ms\ '\kappa)\rightarrow(mg\ '\alpha)))$

for every sentence $'\alpha$ and every set of sentences $'\kappa$ of that FOL object language. The necessity symbol is represented by a box: []. C1 states that a sentence of FOL is a FOL-theorem (i.e. fol) of a set of sentences of FOL if and only if the meaning of that set of sentences necessarily implies the meaning of that sentence. One modal logic which satisfies C1 for FOL is the Z Modal Quantificational Logic described in [Brown 1987; Brown 1989] whose theorems are recursively enumerable. Z has the metatheorem: $(<>\Gamma)\{\pi/\lambda\xi\alpha\}\rightarrow(<>\Gamma)$ where Γ is a sentence of FOL and includes all the laws of S5 Modal Logic [Hughes & Cresswell 1968] whose modal axioms and inference rules are in Figure 6. Therein, κ and Γ rare arbitrary sentences of the intentional semantics.

R0: from α infer $([]\ \kappa)$	A2: $([](\kappa\rightarrow\Gamma))\rightarrow(([]\kappa)\rightarrow([]\Gamma))$
A1: $([]\kappa)\rightarrow\kappa$	A3: $([]\kappa)\vee([]\neg[]\kappa)$

Figure 6: The Laws of S5 Modal Logic

These S5 modal laws and the laws of FOL given in Figure 4 constitute an S5 Modal Quantificational Logic similar to [Carnap 1946; Carnap 1956], and a FOL version [Parks 1976] of [Bressan 1972] in which the Barcan formula: $(\forall\gamma([]\kappa))\rightarrow([]\forall\gamma\kappa)$ and its converse hold. The R0 inference rule implies that anything derivable in the metatheory is necessary. Thus, in any logic with R0, contingent facts would never be asserted as additional axioms of the metatheory. The defined Modal symbols are in Figure 7 with their definitions and interpretations.

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$<>\kappa$	$\neg\ []\ \neg\kappa$	α is logically possible		$[\kappa]\ \Gamma$	$[]\ (\kappa\rightarrow\Gamma)$	β entails α
$\kappa\equiv\Gamma$	$[]\ (\kappa\leftrightarrow\Gamma)$	α is logically equivalent to β		$<\kappa>\ \Gamma$	$<>\ (\kappa\wedge\Gamma)$	α and β is logically possible

Figure 7: Defined Symbols of Modal Logic

From the laws of the Intensional Semantics we prove that the meaning of the set of FOL consequences of a set of sentences is the meaning of that set of sentences (C2), the FOL consequences of a set of sentences contain the FOL consequences of another set if and only if the meaning of the first set entails the meaning of the second set (C3), and the sets of FOL consequences of two sets of sentences are equal if and only if the meanings of the two sets are logically equivalent (C4):

C2: $(ms(fol\ '\kappa))\equiv(ms\ '\kappa)$ proof: The proof divides into two cases:

(1) $[(ms\ '\kappa)](ms(fol\ '\kappa)))$ Unfolding the second ms gives: $[(ms\ '\kappa)]\forall s((s\varepsilon(fol\ '\kappa))\rightarrow(mg\ s))$

By the soundness part of C1 this is equivalent to: $[(ms\ '\kappa)]\forall s(([(ms\ '\kappa)](mg\ s))\rightarrow(mg\ s))$

By the S5 laws this is equivalent to: $\forall s(([(ms\ '\kappa)](mg\ s))\rightarrow [(ms\ '\kappa)](mg\ s))$ which is a tautology.

(2) $[(ms(fol\ '\kappa))](ms\ '\kappa)$ Unfolding ms twice gives: $[\forall s((s\varepsilon(fol\ '\kappa))\rightarrow(mg\ s))]\forall s((s\varepsilon'\kappa)\rightarrow(mg\ s))$

which is: $[\forall s((s\varepsilon(fol\ '\kappa))\rightarrow(mg\ s))]((s\varepsilon'\kappa)\rightarrow(mg\ s))$ Backchaining on the hypothesis and then dropping it gives: $(s\varepsilon'\kappa)\rightarrow(s\varepsilon(fol\ '\kappa))$. Folding $\supseteq$ gives an instance of FOL1. QED.

C3: (fol 'κ)⊒(fol 'Γ) ↔ ([(ms 'κ)](ms 'Γ)) proof: Unfolding ⊒ gives: ∀s((sε(fol 'Γ))→(sε(fol 'κ)))

By C1 twice this is equivalent to: ∀s(([(ms 'Γ)](mg s))→([(ms 'κ)](mg s)))

By the laws of S5 modal logic this is equivalent to: ([(ms 'κ)]∀s(([(ms 'Γ)](mg s)))→(mg s)))

By C1 this is equivalent to: [(ms 'κ)]∀s((sε(fol 'Γ))→(mg s)). Folding ms then gives: [(ms 'κ)](ms(fol 'Γ))

By C2 this is equivalent to: [(ms 'κ)](ms 'Γ). QED.

C4: ((fol 'κ)=(fol 'Γ)) ↔ ((ms 'κ)≡(ms 'Γ)) proof: This is equivalent to (((fol 'κ)⊒(fol 'Γ))∧((fol 'Γ)⊒(fol 'κ))) ↔ ([(ms 'κ)](ms 'Γ))∧([(ms 'Γ)](ms 'κ)) which follows by using C3 twice.

5. "Recursive" Default Logic Represented in Modal Logic

The fixed point equation for Default Logic may be expressed as a necessary equivalence in an S5 Modal Quantificational Logic using a recursive definition, as follows:

$$\kappa \equiv (DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$$

where DR is defined as:

$(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i) =df\ \forall t(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$

$(R\ 0\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i) =df\ \Gamma$

$(R\ t{+}1\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i) =df\ (R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i) \wedge \forall_i((([(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge\wedge_{j=1,mi}(<\kappa>\beta_{ij}))\rightarrow\chi_i)$

When the context is obvious $\Gamma\ \alpha_i{:}\beta_{ij}/\chi_i$ is omitted and just (DR κ) and (R t κ) are written.Given below are some properties of DR. The first two theorems state that DR entails Γ and any conclusion χ_i of a default whose entailment condition holds in DL and whose possible conditions are possible with κ.

MD1: $[(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\Gamma$ proof: Unfolding DR gives:$[\forall t(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\Gamma$

Letting t be 0 shows that it suffices to prove: $[(R\ 0\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\Gamma$. Unfolding R gives a tautology. QED.

MD2:$([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\ (([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge\wedge_{j=1,mi}(<\kappa>\beta_{ij}))\rightarrow \chi_i)$

proof: By R0 it suffices to prove: $(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)\rightarrow ((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge_{j=1,m_i}(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Unfolding DR gives: $\forall t(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i)\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge_{j=1,m_i}(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Letting the quantified t be t+1, it suffices to prove:

$(R\ t{+}1\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i)\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge_{j=1,m_i}(<\kappa>\beta_{ij})))\rightarrow\chi_i)$. Unfolding R t+1 gives:

$((R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))$

$\wedge(\forall_i((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i)]\alpha_i)\wedge(\wedge_{j=1,m_i}(<\kappa>\beta_{ij})))\rightarrow\chi_i)))\wedge([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge_{j=1,m_i}(<\kappa>\beta_{ij})))\rightarrow\chi_i$

Letting the quantified i be i gives a tautology. QED.

The concept (i.e. ss) of the combined meaning of all the sentences of the FOL object language whose meanings are entailed by a proposition is defined as follows:

$$(ss\ \kappa)\ =df\ \forall s(([\kappa](mg\ s))\rightarrow(mg\ s))$$

SS1 shows that a proposition entails the combined meaning of the FOL object language sentences that it entails. SS2 shows that if a proposition is necessarily equivalent to the combined meaning of the FOL object language sentences that it entails, then there exists a set of FOL object language sentences whose meaning is necessarily equivalent to that proposition:

SS1: [κ](ss κ) proof: By R0 it suffices to prove: κ→(ss κ). Unfolding ss gives: κ→∀s(([κ](mg s))→(mg s)) which is equivalent to:∀s(([κ](mg s))→(κ→(mg s))) which is an instance of A1. QED.

SS2: (κ≡(ss κ))→ ∃s(κ≡(ms s)) proof: Letting s be {s: ([κ](mg s)) gives: (κ≡(ss κ))→ (κ≡(ms{s: ([κ](mg s)))). Unfolding ms and lambda conversion gives: (κ≡(ss κ))↔ (κ≡∀s(([κ](mg s))→(mg s))). Folding ss gives a tautology. QED.

The theorems MD3 and MD4 are analogous to MD1 and MD2 except that DR is replaced by the combined meanings of the sentences entailed by DR.

MD3: $[ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)]\forall i\Gamma_i$ proof: By R0 it suffices to prove: $(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))\rightarrow\forall i\Gamma$ which is equivalent to: $(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))\rightarrow\Gamma_i$ Unfolding ss gives: $\forall s(([(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](mg\ s))\rightarrow(mg\ s)))\rightarrow\Gamma_i$ which by the meaning laws M0-M7 is equivalent to: $(\forall s(([(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](mg\ s))\rightarrow(mg\ s)))\rightarrow(mg\ '\Gamma_i)$. Backchaining on $(mg\ '\Gamma_i)$ with s in the hypothesis being $'\Gamma_i$ in the conclusion shows that it suffices to prove: $([(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](mg\ '\Gamma_i)$ which by the meaning laws: M0-M7 is equivalent to: $([(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)]\Gamma_i)$ which by S5 Modal Logic is equivalent to: $([(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)]\forall i\Gamma_i)$ which is an instance of theorem MD1. QED.

MD4: $[ss(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\ ((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

proof: By R0 it suffices to prove: $(ss(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/(mg\ '\chi_i))]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Unfolding ss: $(\forall s(([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)](mg\ s))\rightarrow(mg\ s))\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i))]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Instantiating s in the hypothesis to $'\chi_i$ and then dropping the hypothesis gives:

$(([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)](mg\ '\chi_i))\rightarrow(mg\ '\chi_i))\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i))]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Using the meaning laws M0-M7 gives:

$(([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\chi_i)\rightarrow\chi_i)\rightarrow((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\ \chi_i))]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$

Backchaining on χ_i shows that it suffices to prove:

$((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/(mg\ '\chi_i))](\chi_i)$

By the laws of S5 modal logic this is equivalent to:

$([(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\ ((([(R\ t\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge(\wedge j{=}1,m_i(<\kappa>\beta_{ij})))\rightarrow\chi_i)$ which is MD2. QED.

Theorems MD5 and MD6 show that R is entailed by the meanings of the sentences entailed by DR:

MD5: $[ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](R\ 0\ k(\forall i\Gamma_i))\alpha_i{:}\beta_{ij}/\chi_i))$

proof: Unfolding R 0 gives: $(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))\rightarrow(\forall i\Gamma_i)$ which holds by MD3. QED.

MD6: $([ss(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)](R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))\rightarrow([ss(DR\ \kappa\ \alpha_i{:}\beta_{ij}/\chi_i)](R\ t{+}1\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))$

proof: Unfolding R t+1 in the conclusion gives:

$([ss(DR\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]((R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)\wedge\forall_i((([(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge\wedge j{=}1,m_i(<\kappa>\beta_{ij}))\rightarrow(\chi_i)))$

Using the hypothesis gives: $([ss(DR\ \kappa\ \alpha_i{:}\beta_{ij}/\chi_i]\ \forall_i((([(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge\wedge j{=}1,m_i(<\kappa>\beta_{ij}))\rightarrow\chi_i)$

which holds by MD4. QED.

Finally MD7 and MD8 show that talking about the meanings of sets of FOL sentences in the modal representation of Default Logic is equivalent to talking about propositions in general.

MD7: $(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))\equiv(DR\ \kappa(\forall i\Gamma_i)\ \alpha_i{:}\beta_{ij}/\chi_i)$

proof: In view of SS1, it suffices to prove: $([ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$

Unfolding the second occurrence of DR gives: $[ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)]\forall t(R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$

which is equivalent to: $\forall t([ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))$

By induction on t the proof divides into a base case and an induction step:

(1)Base Case: $([ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](R\ 0\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))$ which holds by theorem MD5.

(2)Induction Step: $([ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](R\ t\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))\rightarrow([ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)](R\ t{+}1\ k\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i))$

which holds by theorem MD6. QED.

MD8: $(\kappa\equiv(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))\rightarrow\exists s(\kappa\equiv(ms\ s))$ proof: From the hypothesis and MD7

$\kappa\equiv(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))$ is derived. Using the hypothesis to replace $(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$ by κ in this result gives: $\kappa\equiv(ss(DR\ \kappa(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))$, By SS2 this implies the conclusion. QED.

6. Conclusion: The Relationship between "Recursive" Default Logic and the Modal Logic

The relationship between the "recursive" set theoretic definition of Default Logic [Reiter 1980] and the modal representation of it is proven in two steps. First theorem R1 shows that the meaning of the set r is the proposition R. Theorem DR1 shows that the meaning of the set dr is the proposition DR. DL2 shows that a set of FOL sentences which contains its FOL theorems is a fixedpoint of the fixedpoint equation of Default Logic with an initial set of axioms and defaults if and only if the meaning (or rather disquotation) of that set of sentences is logically equivalent to DR of the meanings of that initial set of sentences and those defaults.

R1: $(ms(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\equiv(R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$

proof: Inducting on the numeric variable t gives a base case and an induction step:

(1) The Base Case: $(ms(r\ 0(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\equiv(R\ 0(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$. Starting from $(ms(r\ 0(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$ unfolding r gives: $(ms(fol\{'\Gamma i\}))$. By C2 this is equivalent to: $(ms\{'\Gamma i\})$. By MS2 this is equivalent to: $(\forall i\Gamma_i)$. Folding R then gives: $(R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$ which proves the base case.

(2) The Induction Step: $((ms(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\equiv(R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))$

$$\rightarrow((ms(r\ t{+}1(fol\ '\kappa)'\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\equiv(R\ t{+}1(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i))$$

Setting aside the induction hypothesis, we start from: $(ms(r\ t{+}1(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$

Unfolding r gives: $(ms(fol((r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\})))$

By C2 this is equivalent to: $(ms((r\ t\ '\kappa\{'\Gamma i\}\ '\alpha_i{:}'\beta_{ij}/'\chi_i)\cup\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\})))$

By MS3 this is equivalent to: $((ms(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge(ms\{'\chi_i{:}\ ('\alpha_i\varepsilon(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa)\}))$

By MS2 this is : $((ms(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\forall_i((('\alpha_i\varepsilon(r\ t\ '\kappa\{'\Gamma i\}\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\wedge_{j=1,mi}('(\neg\beta_{ij})\notin'\kappa))\rightarrow(mg\ '\chi_i))$

Using C1 twice gives and folding <κ> gives:

$((ms(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\forall_i((([ms(r\ t\ '\kappa\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i)](mg\ '\alpha_i))\wedge\wedge_{j=1,mi}(<(ms\ '\kappa)>(mg\ '\beta_{ij})))\rightarrow(mg\ '\chi_i))$

Using the M0-M7 gives: $((ms(r\ t\ '\kappa\{'\Gamma i\}''\alpha_i{:}'\beta_{ij}/'\chi_i))\wedge\forall_i((([ms(r\ t\ '\kappa\{'\Gamma i\}''\alpha_i{:}'\beta_{ij}/'\chi_i)]\alpha_i)\wedge\wedge_{j=1,mi}(<(ms\ '\kappa)>\beta_{ij}))\rightarrow\chi_i)$

Using the induction hypothesis twice gives:

$((R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)\wedge\forall_i((([(R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)]\alpha_i)\wedge\wedge_{j=1,mi}(<(ms\ '\kappa)>\beta_{ij}))\rightarrow\chi_i)$

Folding R then gives: $((R\ t{+}1(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$ which proves the Induction Step. QED.

DR1: $(ms(dr(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\equiv(DR(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$

proof: $(ms(dr(fol\ '\kappa)'\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$ Unfolding the definition of dr gives: $ms(\cup_{t=1,\omega}(r\ t(fol\ '\kappa)'\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$

Unfolding $\cup$ gives: $ms\{s{:}\ \exists t(s\varepsilon(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$. Unfolding ms gives: $\forall s((s\varepsilon\{s{:}\ \exists t(s\varepsilon(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\})\rightarrow(mg\ s))$ which is equivalent to: $\forall s((\exists t(s\varepsilon(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i)))\rightarrow(mg\ s))$ which is equivalent to: $\forall t\forall s((s\varepsilon(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))\rightarrow(mg\ s))$. Folding ms gives: $\forall t(ms(r\ t(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$

By R1 this is equivalent to: $\forall t(R\ t(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$. Folding DR then gives $(DR(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$ QED.

DR2: $((fol\ '\kappa)=(dr(fol\ '\kappa)\{'\Gamma i\}\ '\alpha_i{:}'\beta_{ij}/'\chi_i))\leftrightarrow((ms\ '\kappa)\equiv(DR(ms\ '\kappa)(ms\ '\Gamma)\alpha_i{:}\beta_{ij}/\chi_i)))$

proof: By FOL9 $(fol\ '\kappa)=(dr(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$ is: $(fol\ '\kappa)=(fol(dl(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i)$. By C4 this is equivalent to: $(ms\ '\kappa)\equiv(ms(dr(fol\ '\kappa)\{'\Gamma i\}'\alpha_i{:}'\beta_{ij}/'\chi_i))$. By DR1 this is equivalent to: $(ms\ '\kappa)\equiv(DR(ms\ '\kappa)(\forall i\Gamma_i)\alpha_i{:}\beta_{ij}/\chi_i)$ QED.

Theorem DR2 shows that the set of theorems: $(fol\ '\kappa)$ of a set $'\kappa$ is a fixedpoint of a fixed point equation of Default Logic if and only if the meaning $(ms\ '\kappa)$ of $'\kappa$ is a solution to the necessary equivalence. Furthermore, by FOL10 there are no other fixedpoints (such as a set not containing all its theorems) and by MD8 there are no other solutions (such as a proposition not representable as a sentence in the FOL object language). Therefore, the Modal representation of Default Logic (i.e. DR), faithfully represents the set theoretic description of the "recursive" definition of Default Logic (i.e. dr). Finally, we note that ($\forall i \Gamma_i$) and (ms 'κ) may be generalized to be arbitrary propositions Γ and κ giving the more general modal representation: $\kappa \equiv$(DR κ Γ α_i:β_{ij}/χ_i).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Numbers: MIP-9526532, EIA-9818341, and EIA-9972843.

Bibliography

[Anatoniou 1997] Antoniou, Grigoris 1997. *NonMonotonic Reasoning*, MIT Press.

[Bressan 1972] Bressan, Aldo 1972. *A General Interpreted Modal Calculus,* Yale University Press.

[Boyer&Moore 1981] R. S. Boyer and J. Strother Moore, "Metafunctions: proving them correct and using them efficiently as new proof procedures," *The Correctness Problem in Computer Science,* R. S. Boyer and J. Strother Moore, eds., Academic Press, New York, 1981.

[Brown 1986] Brown, Frank M. 1986. "Reasoning in a Hierarchy of Deontic Defaults", *Proceedings of the Canadian Artificial Intelligence Conference CSCI 86*, Montreal, Canada, Morgan-Kaufmann, Los Altos.

[Brown 1987] Brown, Frank M. 1987. "The Modal Logic Z", In *The Frame Problem in AI*"*; Proc. of the 1987 AAAI Workshop*, Morgan Kaufmann, Los Altos, CA .

[Brown 1989] Brown, Frank M. 1989. "The Modal Quantificational Logic Z Applied to the Frame Problem", advanced paper *First International Workshop on Human & Machine Cognition*, May 1989 Pensacola, Florida. Abreviated version published in *International Journal of Expert Systems Research and Applications, Special Issue: The Frame Problem. Part A.* eds. Keneth Ford and Pattrick Hayes, vol. 3 number 3, pp169-206 JAI Press 1990. Reprinted in *Reasoning Agents in a Dynamic World: The Frame problem*, editors: Kenneth M. Ford, Patrick J. Hayes, JAI Press 1991.

[Carnap 1946] Carnap, Rudolf 1946. "Modalities and Quantification" *Journal of Symbolic Logic*, vol. 11, number 2.

[Carnap 1956] Carnap, Rudolf 1956. *Meaning and Necessity: A Study in the Semantics of Modal Logic*, The University of Chicago Press.

[Fine 1970] Fine, K. 1970. "Propositional Quantifiers in Modal Logic" *Theoria* 36, p336--346.

[Hughes & Cresswell 1968] Hughes, G. E. and Cresswell, M. J., 1968. *An Introduction to Modal Logic*, Methuen & Co. Ltd., London.

[Leasure 1993] Leasure, David E., 1993. The Modal Logic Z Applied to Lifschitz's Benchmark problems for Formal Nonmonotonic Reasoning, University of Kansas Dissertation, University of Kansas Library.

[Mendelson 1964] Mendelson, E. 1964. *Introduction to Mathematical Logic*, Van Norstrand, Reinhold Co., New York.

[Parks 1976] Parks, Z. 1976. "An Investigation into Quantified Modal Logic", Studia Logica 35, p109-125.

[Quine 1969] Quine, W.V.O., Set Theory and Its Logic, revised edition, Oxford University Press, London, 1969.

[Reiter 1980] Reiter, R. 1980. "A Logic for Default Reasoning" *Artificial Intelligence*, 13.

Author information

Frank M. Brown- University of Kansas, Lawrence, Kansas, 66045, e-mail: brown@ku.edu.

ON THE RELATIONSHIPS AMONG QUANTIFIED AUTOEPISTEMIC LOGIC, ITS KERNEL, AND QUANTIFIED REFLECTIVE LOGIC

Frank M. Brown

***Abstract**: A Quantified Autoepistemic Logic is axiomatized in a monotonic Modal Quantificational Logic whose modal laws are slightly stronger than S5. This Quantified Autoepistemic Logic obeys all the laws of First Order Logic and its L predicate obeys the laws of S5 Modal Logic in every fixed=point. It is proven that this Logic has a kernel not containing L such that L holds for a sentence if and only if that sentence is in the kernel. This result is important because it shows that is superfluous thereby allowing the original equivalence to be simplified by eliminating L from it. It is also shown that the Kernel of Quantified Autoepistemic Logic is a generalization of Quantified Reflective Logic, which coincides with it in the propositional case.*

***Keywords**: Quantified Autoepistemic Logic, Quantified Reflective Logic, Modal Logic, Nonmonotonic Logic.*

1. Introduction

Quantified Autoepistemic Logic (i.e. QAEL) is a generalization of Autoepistemic Logic [Moore, Konolige87, Konolige87b], where both universally and existentially quantified variables are allowed to cross the scope of the L predicate. In a recent paper [Brown 2003b, 2004d] showed that Autoepistemic Logic could be represented in an extension of S5 Modal Logic. This modal representation may be generalized to provide a Quantified Autoepistemic Logic with the following necessary equivalence:

$$\kappa \equiv (\text{QAEL } \kappa\ \Gamma)$$

where QAEL is defined as follows:

$(\text{QAEL } \kappa\ \Gamma) =\text{df } \Gamma \wedge \forall i \forall \xi_i((L\ '\chi_i) \leftrightarrow ([\kappa]\chi_i))$

$(L\ '\chi_i) =\text{df } (L'\chi_i\ a_i)$,

where χ_i is the ith sentence with or without free variables of a First Order Logic (i.e. FOL) and a_i is an association list associating the free variables in $'\chi_i$ to values specified by the sequence of variables ξ_i. The $\forall i$ quantifier ranges across the natural numbers. This Quantified Autoepistemic Logic is important because unlike some other attempts [Konolidge1989] to generalize Autoepistemic Logic, its quantifiers obey both the Barcan Formulae, the converse of the Barcan formula, and also all the laws of S5 Modal Logic and First Order Logic (i.e. FOL). Interpreted doxastically this necessary equivalence states that:

that which is believed is equivalent to: Γ and for all i and for all ξ_i $(L\ '\chi_i)$ if and only if χ_i is believed.

The purpose of this paper is to show that the L predicate is not essential to solving for κ and can be eliminated thereby allowing the above necessary equivalence to be replaced by a simpler necessary equivalence which when interpreted as a doxastic logic states:

that which is believed is equivalent to: Γ (with each *L'* replaced by $[\kappa]$) .

thereby eliminating every occurrence of the L predicate, all the (quoted) names of sentences, and the bi-implication containing *L*.

The remainder of this paper proves that the L predicate can be eliminated. Section 2 describes the First Order Logic (i.e.FOL) used herein. Section 3 describes the Modal Logic used herein. QAEL is defined in more detail in section 4. The *L* eliminated form of Quantified Autoepistemic Logic herein called the Quantified Autoepistemic Kernel (i.e. QAEK) is defined in section 5 and is explicated with theorems LEXT1 and LEXT2. In section 6, QAEK is shown to be related to QAEL by theorems QAEK1, QAEK2, QAEK3. The relationship between QAEK and Quantified Reflective Logic (i.e. QRL) [Brown 2003a] is given in section 7. Finally, in section 8, some consequences of all these results are discussed. Figure 1 outlines the relationship of all these theorems in producing the final theorems LEXT2, QAEK3, and AR2.

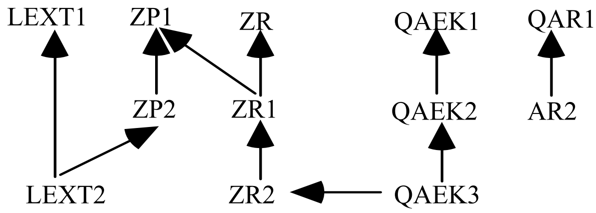


Figure 1: Dependencies among the Theorems

2. First Order Logic

We use a First Order Logic (i.e. FOL) defined as the six tuple: ($\rightarrow$, #f, $\forall$, *vars*, *predicates*, *functions*) where $\rightarrow$, #f, and $\forall$ are logical symbols, *vars* is a set of variable symbols, *predicates* is a set of predicate symbols each of which has an implicit arity specifying the number of associated terms, and *functions* is a set of function symbols each of which has an implicit arity specifying the number of associated terms. Lower case Roman letters possibly indexed with digits are used as variables. Greek letters possibly indexed with digits or lower case roman letters are used as syntactic metavariables. γ, $\gamma_1 ... \gamma_n$, range over the variables, ξ, $\xi_1 ... \xi_n$ range over sequences of variables of an appropriate arity, $\pi, \pi_1 ... \pi_n$ range over the predicate symbols, ϕ, $\phi_1 ... \phi_n$ range over function symbols, δ, $\delta_1 ... \delta_n$, σ range over terms, and α, $\alpha_1 ... \alpha_n$, β, $\beta_1 ... \beta_n, \chi$, $\chi_1 ... \chi_n, \Gamma, \Gamma_1, ... \Gamma_n, \kappa, \kappa_1, ... \kappa_n$, φ, range over sentences (including sentences with free variables). The terms are of the forms γ and (ϕ $\delta 1 ... \delta n$), and the sentences are of the forms ($\alpha \rightarrow \beta$), #f, ($\forall \gamma$ α), (π $\delta_1 ... \delta_n$), and ($\delta = \sigma$). A nullary predicate π or function ϕ is written without parentheses. $\varphi\{\pi / \lambda \xi \alpha\}$ represents the replacement of all occurrences of π in φ by α followed by lambda conversion. The primitive symbols are shown in Figure 2 with their interpretations. The particular FOL used herein includes the binary predicate symbol L and a denumerably infinite number of 0-ary function symbols representing the names (i.e. 'α) of the sentences (i.e. α) of this FOL.

Symbol	Meaning
$\alpha \rightarrow \beta$	if α then β.
#f	falsity
$\forall \gamma$ α	for all γ, α.

Figure 2: Primitive Symbols of First Order Logic

The defined symbols are listed in Figure 3 with their definitions and intuitive interpretations.

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$\neg\alpha$	$\alpha \rightarrow$ #f	not α		$\alpha \wedge \beta$	$\neg(\alpha \rightarrow \neg\beta)$	α and β
#t	$\neg$ #f	truth		$\alpha \leftrightarrow \beta$	$(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$	α if and only if β
$\alpha \vee \beta$	$(\neg \alpha) \rightarrow \beta$	α or β		$\exists \gamma$ α	$\neg \forall \gamma \neg \alpha$	for some γ , α

Figure 3: Defined Symbols of First Order Logic

3. Modal Logic

The We extend First Order Logic with a necessity symbol as given in Figure 4 below:

Symbol	Meaning
[]α	α is logically necessary

Figure 4: Primitive Symbols of Modal Logic

and with the laws of an S5 Modal Logic [Hughes & Cresswell 1968] as given in Figure 5 below:

R0: from α infer ([] α)	A2: ([]($\alpha \rightarrow \beta$)) $\rightarrow$ (([]α) $\rightarrow$ ([]β))
A1: ([]α) $\rightarrow$ α	A3: ([]α) $\vee$ ([]$\neg$[]α)

Figure 5: The Laws of S5 Modal Logic

These S5 modal laws and the laws of FOL constitute an S5 Modal Quantificational Logic similar to [Carnap 1946; Carnap 1956], and a FOL version [Parks 1976] of [Bressan 1972] in which the Barcan formula: $(\forall\gamma([]\alpha))\rightarrow([]\forall\gamma\alpha)$ and its converse hold. The defined Modal symbols used herein are listed in Figure 6 with their definitions and intuitive interpretations.

Symbol	Definition	Meaning		Symbol	Definition	Meaning
$<> \alpha$	$\neg [] \neg\alpha$	α is logically possible		$\alpha\equiv \beta$	$[] (\alpha\leftrightarrow\beta)$	α is synonymous to β
$<\beta> \alpha$	$<> (\beta\wedge\alpha)$	α and β is logically possible		$\delta_1=\delta_2$	$(\pi\delta_1)\equiv(\pi\ \delta_1)$	δ_1 is logically equal to δ_2
$[\beta]\ \alpha$	$[]\ (\beta\rightarrow\alpha)$	β entails α		$\delta\neq\sigma$	$\neg(\delta=\sigma)$	δ is not logically equal to σ.

Figure 3: Defined Symbols of Modal Logic

Next, we extend the FOL + S5 Modal Quantificational Logic with the A4 axiom scheme given in Figure 7.

A4: $<>\Gamma\{\pi/\lambda\xi\alpha\} \rightarrow <>\Gamma$

where $\Gamma\{\pi/\lambda\xi\alpha\}$ is the simultaneous replacement in Γ of all unmodalized occurrences of π by α.

Figure 7: The Possibility Axiom Scheme

Intuitively, A4 specifies that a sentence Γ is logically possible whenever the result obtained by "interpreting" all the unmodalized occurrences of a predicate within it, is logically possible. If A4 is successively applied to all the unmodalized predicates then it follows that a sentence Γ is logically possible if the result of interpreting all the unmodalized predicates is logically possible. The possibility axiom A4 extends the trivial possibility axiom (i.e. some proposition is neither #t nor #f) given in [Lewis 1936] and [Bressan 1972], the S5c possibility axiom schema (i.e. every conjunction of distinct negated or unnegated propositional constants is logically possible) given in [Hendry & Pokriefka 1985], and is implied by the possibility axiom schema used in the Z Modal Quantificational Logic described in [Brown 1987; Brown 1989]. The following metatheorems are derivable:

ZP1: The Possibility of a Separable Predicates: If (1) Γ, α, and β are sentences of FOL extended whereby any modalized sentence may occur in the place of predicates and (2) π does not occur unmodalized in any of Γ, α, and β then: $(<>(\Gamma\wedge(\forall\xi(\alpha\rightarrow(\pi\ \xi)))\wedge(\forall\xi(\beta\rightarrow\neg(\pi\ \xi)))))\leftrightarrow(<>(\Gamma\wedge(\neg\exists\xi(\alpha\wedge\beta))))$

ZP2: The Possibility of a Defined Predicate: If (1) Γ and α are sentences of FOL without equality but extended whereby any modalized sentence may occur in the place of predicates and (2) π does not occur unmodalized in any of Γ and α then: $(<>(\Gamma\wedge(\forall\xi((\pi\ \xi)\leftrightarrow\alpha))))\leftrightarrow(<>\Gamma)$ proof: Let β be $\neg\alpha$ in ZP1 and simplify. QED.

ZR: The Reduction Lemma: If (1) κ occurs in Γ and Ψ only in the context: $<\kappa>\varphi$ for some φ (or in the context $[\kappa]\mu$ which is essentially of the same modal form: $\neg<\kappa>\neg\mu$) and (2) for all such φ: $\forall p(((<(\Gamma\wedge\Psi)>\varphi)\leftrightarrow(<(\Gamma>\varphi))\{\kappa/p\})$ then : $(\kappa\equiv(\Gamma\wedge\Psi)) \leftrightarrow \exists p((\kappa\equiv(p\wedge(\Psi\{\kappa/p\})))\wedge(p\equiv(\Gamma\{\kappa/p\})))$

ZR1: Reducing a Reflection with a Separable Predicate: If (1) κ occurs in Γ, α, and β only in the context: $<\kappa>\varphi$ for some φ (or in the context $[\kappa]\mu$ which is essentially of the same modal form: $(\neg<\kappa>\neg\mu)$, (2) Γ, α, β, and φ are sentences of FOL extended whereby any modalized sentence may occur in the place of predicates, (3) π does not occur unmodalized in any of Γ, α, β, and φ then: $(\kappa\equiv(\Gamma\wedge(\forall\xi(\alpha\rightarrow(\pi\ \xi)))\wedge(\forall\xi(\beta\rightarrow\neg(\pi\ \xi))))$

$\leftrightarrow \exists p((\kappa\equiv(p\wedge(((\forall\xi(\alpha\rightarrow(\pi\ \xi)))\wedge(\forall\xi(\beta\rightarrow\neg(\pi\ \xi)))))\{\kappa/p\})))\wedge(p\equiv((\Gamma\wedge(\neg\exists\xi(\alpha\wedge\beta)))\{\kappa/p\})))$

ZR2: Reducing a Reflection with a Defined Predicate: If (1) κ occurs in Γ, and α only in the context: $<\kappa>\varphi$ for some φ (or in the context $[\kappa]\mu$ which is essentially of the same modal form: $(\neg<\kappa>\neg\mu)$, (2) Γ, α, , and φ are sentences of FOL extended whereby any modalized sentence may occur in the place of predicates, (3) π does not occur unmodalized in any of Γ, α, and φ then: $\kappa\equiv(\Gamma\wedge\forall\xi((\pi\ \xi)\leftrightarrow\alpha))\leftrightarrow \exists p\ \kappa\equiv(p\wedge\forall\xi((\pi\ \xi)\leftrightarrow\alpha\{\kappa/p\})) \wedge p\equiv\Gamma\{\kappa/p\}$. proof: Let β be $\neg\alpha$ in ZR1 and simplify. QED.

4. Quantified Autoepistemic Logic

Quantified Autoepistemic Logic (i.e. QAEL) is defined in Modal Logic by a necessary equivalence of the form:

$$\kappa\equiv(\text{QAEL}\ \kappa\ \Gamma)$$

where QAEL is defined as follows: $(\text{QAEL}\ \kappa\ \Gamma) =\text{df}\ \Gamma\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$ where $(L\ '\chi_i) =\text{df}\ (L\ '\chi_i\ a_i)$, χ_i is the ith sentence with or without free variables of FOL and a_i is an association list binding the free variables in $'\chi_i$ to values specified by the sequence of metalanguage variables ξ_i. The $\forall$i quantifier ranges across the natural numbers. Any FOL proposition κ which makes this necessary equivalence true is a solution. QAEL addresses the problem of how quantified variables whose scopes cross the L predicate may be represented. Furthermore these quantifiers obey not only the Barcan formula but unlike the generalization of Autoepistemic Logic given in [Konolige 1989] its converse and therefore does not suffer the anomalies therein discussed. For example we could then state in QAEL that everything is a bird and that all things believed to be birds for which flying is believable do in fact fly as follows: $\kappa\equiv$(QAEL κ(($\forall$x(Bird x))$\wedge\forall$x(((L '(Bird x))$\wedge$($\neg L$ '($\neg$(Fly x))))$\rightarrow$(Fly x))))

5. Quantified Autoepistemic Kernel

The Quantified Autoepistemic Kernel [Brown 1989] is defined in Modal Quantificational Logic by the necessary equivalence:

$$\varphi\equiv(\text{QAEK } \varphi\ \Gamma)$$

where QAEK is defined as: (QAEK φ Γ) =df $\Gamma\{L\ '/[\varphi]\}$

The L predicate does not occur in QAEK. However, the kernel may be used to define an extension containing facts involving L as follows:

(L-EXT φ) =df ($\varphi\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow(([\varphi]\chi_i)\{L\ '/[\varphi]\})))$

The kernel φ possesses two important properties with respect to L-extensions, namely that the L-extension of φ entails φ, and φ entails every FOL sentence not containing an occurrence of L which the L-extension entails

LEXT1: [(L-ext φ)]φ proof: Unfolding L-ext gives a tautology. QED.

LEXT2: If L is not in s and if φ contains no unmodalized occurrence of L, then: $\forall$s([(L-ext φ)]s)$\leftrightarrow$ ([φ]s)

proof: Pushing negation through gives the equivalent sentence: $\forall$r((<(L-ext φ)>r)$\leftrightarrow$ (<φ>r))

Unfolding L-ext gives: $\forall$r((<($\varphi\wedge\forall i\forall\xi_i((L\ \chi_i)\leftrightarrow(([\varphi]\chi_i)\{L\ '/[\varphi]\}))$>r) $\leftrightarrow$ (<φ>r)) or rather:

$\forall$r((<φ>($\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow(([\varphi]\chi_i)\{L\ '/[\varphi]\}))\wedge$r) $\leftrightarrow$ (<φ>r)) which is an instance of theorem ZP2. QED

LEXT1 and LEXT2 show that the kernel determines all the non-kernel sentences in the L-extension. Representing problems in the Quantified Autoepistemic Kernel simplifies their solution since the pre-processing step of eliminating the L predicate from Γ is eliminated.

6. The Relationship between Quantified Autoepistemic Logic and its Kernel

We now show how all occurrences of L including those within quotes as parts of structural descriptive names of sentences of Autoepistemic Logic may be eliminated from Γ: For example, if Γ consisted of the single default: ($\neg$(L '(L '($\neg$ π))))$\rightarrow\pi$ then the necessary equivalence is: $\kappa\equiv$(QAEL κ(($\neg$(L '(L '($\neg$ π))))$\rightarrow\pi$))

Unfolding AEL gives: $\kappa\equiv$((($\neg$(L'(L '($\neg$ π))))$\rightarrow\pi$)$\wedge\forall i\forall\xi_i((L'\chi_i)\leftrightarrow([\kappa]\chi_i))$). Since the quantified statement is connected to: ($\neg$(L '(L '($\neg$ π))))$\rightarrow\pi$ by a conjunction it may be assumed when simplifying that expression. Instantiating i so that χ_i is (L '($\neg$ π)) and using that instance gives the equivalent expression: $\kappa\equiv$((($\neg$([κ](L '($\neg$ π))))$\rightarrow\pi$) $\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$). We would like to eliminate the remaining L in the first formulae but it is inside the scope of an entailment and therefore the (non-necessary) equivalence: $\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$ does not justify such a reduction merely by virtue of the two formulas being connected by conjunction. However, the entire formula allows the derivation of: $[\kappa](\forall i\forall\xi_i((L\ '\chi_i))\leftrightarrow([\kappa]\chi_i))$ which shows that $\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$ may be assumed in any scope entailed by κ. Thus we can still reduce occurrences of L even embedded within an entailment. Thus, the above equation is equivalent to: $\kappa\equiv$((($\neg$([κ]([κ]($\neg$ π))))$\rightarrow\pi$)$\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))$) in which no occurrence of L nor quotation appears in the first formulae in the conjunction. Notating the above described process (i.e. sequence of deductions) as ($\Gamma\{L\ '/[\kappa]\}$) or rather the substitution of L ' by [κ] gives the theorem:

QAEK1: ($\kappa\equiv$(QAEL κ Γ)) $\leftrightarrow$ ($\kappa\equiv$(QAEL κ ($\Gamma\{L\ '/[\kappa]\}$)))

The process which eliminated L from Γ can also be used to eliminate *L* from χ_i in the formulae: $\kappa\equiv((\Gamma\{L\ '/[\kappa]\})\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$. Since χ_i occurs within the modal scope of a κ entailment we are justified in replacing an instance of it by another formulae by assuming $\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$ for any other instance of χ_i since $[\kappa]\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i)))$ follows from the overall equation. To replace each χ_i by a sentence which no longer contains *L*, we specify an ordering of all the sentences based on the maximum depth of *L*s as they occur through the structural descriptive names of the sentences. A sentence χ_i with no *L* would be of depth 0, a sentence with L would be at depth 1, a sentence with 'L would be of depth 2, a sentence with ''*L* would be of depth 3 and so forth. The proof is by induction. The base case is always true since *L* is not in those sentences. The induction step proceeds by using $\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow([\kappa]\chi_i))))$ on sentences whose *L* depth is less than n to prove that relation for sentences whose depth is n.

QAEK2 shows how all but one occurrence of *L* may be eliminated from the equivalence. Essentially κ is logically equivalent to a modal formula $\Gamma\{L\ '/[\kappa]\})$ not containing L conjoined to what is essentially a "definition" of *L* in terms of another modal formulae not containing *L*. This suggests that *L* is superfluous notation and that the essence of κ lies only in the first formulae. This intuition is easily proven:

QAEK3: $(\kappa\equiv(\text{QAEL}\ \kappa\ \Gamma))\leftrightarrow\exists p((\kappa\equiv(\text{L-EXT}\ p))\wedge(p\equiv(\text{QAEK}\ p\ \Gamma)))$

proof: By QAEK2 $(\kappa\equiv(\text{QAEL}\ \kappa\ \Gamma))$ is equivalent to: $(\kappa\equiv((\Gamma\{L\ '/[\kappa]\})\wedge\forall i\forall\xi_i((L\ '\chi_i)\leftrightarrow(([\kappa]\chi_i)\{L\ '/[\kappa]\}))$

Instantiating ZR2 with: $\Gamma:=\Gamma\{L\ '/[\kappa]\}$, $\xi:=i\xi_i$, $\pi:=L$, $\alpha:=([\kappa]\chi_i\{L\ '/[\kappa]\})))$ shows that the above expression is equivalent to: $\exists p((\kappa\equiv(p\wedge\forall i\forall\xi_i((L\ '\ \chi_i)\leftrightarrow(([p]\chi_i)\{L\ \ '/[p]\})))\wedge(p\equiv(\Gamma\{L\ '/[p]\})))$. Folding L-EXT and QAEK gives $\exists p((\kappa\equiv(\text{L-EXT}\ p))\wedge(p\equiv(\text{QAEK}\ p\ \Gamma)))$ QED.

QAEK3 divides the Autoepistemic equation into two distinct equivalences, one axiomatizing the kernel p and the other defining the stronger proposition κ which is the L-extension of p containing additional facts about the *L* predicate. LEXT1 and LEXT2 show that the L-extension κ is a conservative extension of the kernel and therefore it is not essential. For this reason it suffices to deal with just the necessary equivalence for the Quantified Autoepistemic Kernel in studying Quantified Autoepistemic Logic: $\varphi\equiv(\text{QAEK}\ \varphi\ \Gamma)$.

7. The Relationship between Quantified Autoepistemic Kernel and Quantified Reflective Logic

The modal representation of Reflective Logic [Brown 1989, 2003a, 2003c] may be generalized to a Quantified Reflective Logic as:

$$\kappa\equiv(\text{QRL}\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$$

where QRL is defined in Modal Logic as follows:

$(\text{QRL}\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i) =_{df}\ \Gamma\wedge\forall i\forall\xi_i((([\kappa]\alpha_i)\wedge\wedge_{j=1,m_i}\langle\kappa\rangle\beta_{ij})\rightarrow\chi_i))$

where Γ, α_i, β_{ij}, and χ_i are sentences of FOL which may contain free variables. The variables in ξ_i may occur in any of α_i, β_{ij}, and χ_i. When the context is obvious $\Gamma\ \alpha_i{:}\beta_{ij}/\chi_i$ is omitted and instead just (QRL κ) is written. $\wedge_{j=1,m_i}$ stands for the conjunction of the formula which follows it as j ranges from 1 to m_i. If $m_i=0$ then it specifies #t. If i ranges over a finite number of defaults then $\forall_i$ may be replaced in this definition by a conjunction: $\wedge_i$. Interpreted as a doxastic logic, the necessary equivalence states:

that which is believed is logically equivalent to:

Γ and for each i, if α_i is believed and for each j, β_{ij} is believable then χ_i

Quantified Reflective Logic is an instance of the Quantified Autoepistemic Kernel. Specifically:

QAR1: $(\text{QRL}\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)\equiv(\text{QAEK}\ \kappa\ \Gamma\wedge\forall i\forall\xi_i(((L\ '\alpha_i)\wedge\wedge_{j=1,m_i}(\neg(L\ '(\neg\beta_{ij}))))\rightarrow\chi_i))$

proof: Unfolding QRL and QAEK gives identical formulas. QED.

We call the instance of the Quantified Autoepistemic Kernel in which no quantified variables in Γ cross a modal scope simply the Autoepistemic Kernel. (i.e. AEK). Likewise, we call the instance of Quantified Reflective Logic with no variables in the any sequence ξ_i simply Reflective Logic (i.e. RL).

$(AEK\ \varphi\ \Gamma)$ =df $\Gamma\{L\ /[\varphi]\}$

$(RL\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$ =df $\Gamma\wedge\forall_i((([\kappa]\alpha_i)\wedge\wedge_{j=1,mi}\langle\kappa\rangle\beta_{ij})\rightarrow\chi_i)$

where Γ, α_i, β_{ij}, and χ_i are closed sentences of FOL. By closed it is meant that no sentence may contain a free variable.

By QAR1 Reflective Logic is clearly an instance of the Autoepistemic Kernel. However, in addition, it turns out that the Autoepistemic Kernel is also an instance of Reflective Logic:

AR2. The Autoepistemic Kernel is an instance of Reflective Logic. Specifically, for every FOL formulae Γ_i there exist FOL formulas: α_i, β_{ij}, and χ_i. such that: $(AEK\ \kappa\ (\forall_i\Gamma_i))\equiv(RL\ \kappa\ \#t\ \alpha_i{:}\beta_{ij}/\chi_i)$

proof: By QAR1 it suffices to prove that each $\Gamma_i\{L'/[\kappa]\}$, which we shall herebelow call Ψ is representable as

$\wedge_i((([\kappa]\alpha_i)\wedge(\wedge_{j=1,mi}(\langle\kappa\rangle\beta_{ij})))\rightarrow\chi_i)$

We choose a k-entailment: $([\kappa]\varphi)$ in Ψ of lowest scope that has not already been chosen. We use the laws of classical logic to place φ into conjunctive normal form (treating any embedded κ-entailment as another predicate). The following five theorem schemata of Z are then used to reduce the scope of $[\kappa]$[9] .

KU1: $([\kappa](\alpha_1\wedge...\wedge\alpha_n))\equiv(([\kappa]\alpha_1)\wedge...\wedge([\kappa]\alpha_n))$

KU2: $([\kappa](\alpha_1\vee...\vee\alpha_m\vee([\kappa]\varphi)\vee\beta_1\vee...\vee\beta_n))\equiv(([\kappa]\varphi)\vee([\kappa](\alpha_1\vee...\vee\alpha_m\vee\beta_1\vee...\vee\beta_n)))$

KU3: $([\kappa](\alpha_1\vee...\vee\alpha_m\vee(\neg([\kappa]\varphi))\vee\beta_1\vee...\vee\beta_n))\equiv((\neg[\kappa]\varphi)\vee([\kappa](\alpha_1\vee...\vee\alpha_m\vee\beta_1\vee...\vee\beta_n)))$

KU4: $([\kappa]([\kappa]\varphi))\equiv([\kappa]\varphi)$

KU5: $([\kappa](\neg[\kappa]\varphi))\equiv((\neg[\kappa]\varphi)\vee([\kappa]\#f))$

If the result begins with a conjunction, KU1 is applied. If the result begins with a disjunction with an embedded κ entailment or negation of a κ entailment then respectively KU2 or KU3 is applied. If the result is itself a k-entailment of the negation of a k-entailment then respectively KU4 or KU5 is applied. The over all process is repeated until no further KU rule is applicable. When the process finishes since none of the above rules is applicable if the overall formula is put into conjunctive normal form then every resulting disjunction must be of the following form when negations of entailments are ordered before entailments which are ordered before other expressions: $((\vee_{j=1,a}\neg([\kappa]\alpha_j))\vee(\vee_{j=1,b}([\kappa]\beta_j))\vee(\vee_{j=1,c}\chi_j))$ Pulling the first negation out and noting that $(\wedge_{j=1,a}([\kappa]\alpha_j))$ is equivalent to $([\kappa](\wedge_{j=1,a}\alpha_j))$ gives:

$((\neg([\kappa](\wedge_{j=1,a}\alpha_j)))\vee(\vee_{j=1,b}([\kappa]\beta_j))\vee(\vee_{j=1,c}\chi_j))$ or rather: $((\neg([\kappa](\wedge_{j=1,a}\alpha_j)))\vee(\vee_{j=1,b}([\kappa]\beta_j))\vee(\vee_{j=1,c}\chi_j))$

Letting α be $(\wedge_{j=1,a}\alpha_j)$ and χ be $(\vee_{j=1,c}\chi_j)$ gives: $((\neg([\kappa]\alpha))\vee(\vee_{j=1,b}([\kappa]\beta_j))\vee\chi)$

where α is #t if their are no α_j formulas (since that is the identity of conjunction) and where χ is #f if there are no χ_j formulas (since that is the identity of disjunction). Rewriting the above as an implication gives:

$((([\kappa]\alpha)\wedge(\wedge_{j=1,mi}(\langle\kappa\rangle\beta_j)))\rightarrow\chi_i$ where the resulting β_j are the negations of the previous ones. This formula is called a default. The conjunction of all the defaults is then written as:

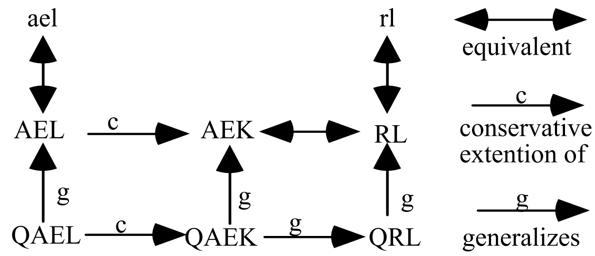
$\wedge_i((([\kappa]\alpha_i)\wedge(\wedge_{j=1,mi}(\langle\kappa\rangle\beta_{ij})))\rightarrow\chi_i)$ where the defaults are not required to have any β subformulas. QED.

$(RL\ \kappa\ \#t\ \alpha_i{:}\beta_{ij}/\chi_i)$ is often written as: $(RL\ \kappa\ \Gamma\ \alpha_i{:}\beta_{ij}/\chi_i)$ where Γ is all those defaults having no α (or where α is #t) nor β subformulas (and hence no modals) and i ranges over just the "real" defaults containing modals.

8 Conclusion

The nonmonotonic systems discussed herein are related as described in Figure 8.

[9]When $([K]\psi)$ is viewed with K fixed as a unary symbol, it has the properties of a KU45 modal logic **[Park]** .



ael = Autoepistemic Logic [Moore 1985]	rl = Reflective Logic [Brown 1986]
AEL = Autoepistemic Logic in Modal Logic	QAEL = Quantified Autoepistemic Logic in Modal Logic
AEK = Autoepistemic Kernel in Modal Logic	QAEK = Quantified Autoepistemic Kernel in Modal Logic
RL = Reflective Logic in Modal Logic	QRL = Quantified Reflective Logic in Modal Logic

Figure 8: Relationships among NonMonotonic Systems

The original set theoretic description of Autoepistemic Logic (i.e. ael) is equivalent [Brown 2003b, 2003d] to the modal description AEL and the set theoretic description of Reflective Logic (i.e. rl) is equivalent [Brown 2003a, 2003c] to the modal description RL. Equivalence means that the meaning of the fixed-points of the set theoretic descriptions are identical to the solutions of the necessary equivalences of the modal systems whenever their inputs bear a similar relation. Since the modal systems (i.e. FOL+S5+A4) are much simpler than the set theoretic descriptions (i.e. FOL + Set Theory + FOL Syntax, + FOL Proof Theory) they provide a reduction in both conceptual and computational complexity. For this reason we focus on the modal systems: AEL and RL.

QAEL and QAEK, are respectively generalizations of AEL and AEK in which quantifiers are allowed to be inserted anywhere in the formulas and where such quantified variables may cross modal scopes. Since AEL and QAEL are proven by QAEK3 and LEXT2 to be conservative extensions (involving the superfluous *L* predicate) of AEK and QAEK respectively, these systems AEK and QAEK are said to be the kernels of AEL and QAEL respectively. Because the kernel systems eliminate all occurrences of *L*' and the biconditional relating *L*' to [k] they are more useful systems for both understanding and automatic theorem proving. For this reason we now focus on just the kernel systems: AEK and QAEK.

AEK is proven to be equivalent to RL by AR2. QRL is a generalization of RL where only universal quantifiers may be inserted and only inserted at the beginning of a default. By QAR1, QAEK is a generalization of QRL, But in general QRL is weaker than QAEK since for example it does not allow for existential quantifiers just before a default. Because QAEK and QRL differ while AEK and RL are equivalent, it follows that both QAEK and QRL can be said to be different quantificational generalizations of the Autoepistemic Kernel. Both are interesting systems with QAEK providing greater generality and QRL having deep relationships to nomonotonic logics with quantified default inference rules [Brown 2003f]. [Brown 2003e] describes an Automatic Deduction system for the propositional case of Autoepistemic Kernels (i.e. AEK) which reduces to the propositional case of Reflective Logic (i.e. RL)). Deduction Methods for the QAEL and QRL are discussed in [Brown 1987; Leasure 1993; Leasure & Brown 1995].

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Numbers: MIP-9526532, EIA-9818341, and EIA-9972843.

Bibliography

[Bressan 1972] Bressan, Aldo 1972. *A General Interpreted Modal Calculus,* Yale University Press.

[Brown 1987] Brown, Frank M. 1987. "The Modal Logic Z", In *The Frame Problem in AI"; Proc. of the 1987 AAAI Workshop*, Morgan Kaufmann, Los Altos, CA .

[Brown 1989] Brown, Frank M. 1989. "The Modal Quantificational Logic Z Applied to the Frame Problem", advanced paper *First International Workshop on Human & Machine Cognition*, May 1989 Pensacola, Florida. Abbreviated version published in *International Journal of Expert Systems Research and Applications, Special Issue: The Frame Problem. Part*

A. eds. Kenneth Ford and Patrick Hayes, vol. 3 number 3, pp169-206 JAI Press 1990. Reprinted in *Reasoning Agents in a Dynamic World: The Frame problem*, editors: Kenneth M. Ford, Patrick J. Hayes, JAI Press 1991.

[Brown 2003a]Frank M. Brown, "On the Relationship between Reflective Logic and Modal Logic", *Proceedings of the X-th International Conference Knowledge Dialogue Solution - Varna 2003*, pages 356 – 364, ISBN 954-16-0025-5, FOI-Commerce, Sofia, 2003.

[Brown 2003b] Frank M, Brown, "On the Relationship between Autoepistemic Logic and Modal Logic", *Proceedings of the X-th International Conference Knowledge Dialogue Solution - Varna 2003*, pages 382-390, ISBN 954-16-0025-5, FOI-Commerce, Sofia 2003.

[Brown 2003c] Frank M. Brown "Representing Reflective Logic in Modal Logic", *Information Theories and Applications*, Vol. 10, no. 4, pages: 431-438, December 2003. (Revised version of "On the Relationship between Reflective Logic and Modal Logic")

[Brown 2003d]Frank M. Brown, "Representing Autoepistemic Logic in Modal Logic", *Information Theories and Applications*, Vol. 10, no. 4, pages 455-462, December 2003. Revised version of "On the Relationship between Autoepistemic Logic and Modal Logic"

[Brown 2003e] Frank M. Brown "Decision Procedures for the Propositional Cases of 2nd Order Logic and Z Modal Logic Representations of a 1st Order L-Predicate Nonmonotonic Logic", *Automated Reasoning with Analytic Tableaux and Related Method: Tableaux 2003*, September 2003, Rome, *Lecture Notes in Artificial Intelligence 2796*, ISBN 3-540-40787-1, Springer-Verlag, Berlin, 2003.

[Carnap 1946] Carnap, Rudolf 1946. "Modalities and Quantification" *Journal of Symbolic Logic*, vol. 11, number 2, 1946.

[Carnap 1956] Carnap, Rudolf 1956. *Meaning and Necessity: A Study in the Semantics of Modal Logic*, The University of Chicago Press.

[Fine 1970] Fine, K. 1970. "Propositional Quantifiers in Modal Logic" *Theoria* 36, p336--346.

[Hendry & Pokriefka 1985] Hendry, Herbert E. and Pokriefka, M. L. 1985. "Carnapian Extensions of S5", *Journal of Phil. Logic 14*.

[Hughes & Cresswell 1968] Hughes, G. E. and Cresswell, M. J., 1968. *An Introduction to Modal Logic*, Methuen & Co. Ltd., London

[Konolige 1987] Konolige, Kurt 1987. "On the Relation Between Default Theories and Autoepistemic Logic", *IJCAI87 Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 1987.

[Konolige 1987b] Konolige, Kurt 1987. "On the Relation Between Default Logic and Autoepistemic Theories", Artificial Intelligence 35(3):343-382.

[Konolige 1989] Konolige, Kurt 1989. "On the Relation between Autoepistemic Logic and Circumscription Preliminary Report", IJCAI89.

[Leasure 1993] Leasure, David E., The Modal Logic Z Applied to Lifschitz's Benchmark problems for Formal Nonmonotonic Reasoning, University of Kansas Ph.D Dissertation, University of Kansas Library , 1993.

[Leasure & Brown 1995] Leasure, David E. and Brown, Frank M. 1995. "AutoEpistemic Logic with Quantifiers", Proceedings Florida Artificial Intelligence Conference.

[Lewis 1936] Lewis, C. I. 1936. Strict Implication, *Journal of Symbolic Logic*, vol I.

[Moore 1985] Moore, R. C. 1985. "Semantical Considerations on Nonmonotonic Logic" *Artificial Intelligence*, 25.

[Park 1988], Park, S., On Formalizing Commonsense Reasoning Using the Modal Situational Logic and Reflective reasoning. Ph.D.dissertation, University of Texas, 1998.

[Parks 1976] Parks, Z. 1976. "An Investigation into Quantified Modal Logic", Studia Logica 35, p109-125.

Author information

Frank M. Brown- University of Kansas, Lawrence, Kansas, 66045, e-mail: brown@ku.edu.

METHODS FOR SOLVING NECESSARY EQUIVALENCES

Frank M. Brown

***Abstract**: Nonmonotonic Logics such as Autoepistemic Logic, Reflective Logic, and Default Logic, are usually defined in terms of set-theoretic fixed-point equations defined over deductively closed sets of sentences of First Order Logic. Such systems may also be represented as necessary equivalences in a Modal Logic stronger than S5 with the added advantage that such representations may be generalized to allow quantified variables crossing modal scopes resulting in a Quantified Autoepistemic Logic, a Quantified Autoepistemic Kernel, a Quantified Reflective Logic, and a Quantified Default Logic. Quantifiers in all these generalizations obey all the normal laws of logic including both the Barcan formula and its converse. Herein, we address the problem of solving some necessary equivalences containing universal quantifiers over modal scopes. Solutions obtained by these methods are then compared to related results obtained in the literature by Circumscription in Second Order Logic since the disjunction of all the solutions of a necessary equivalence containing just normal defaults in these Quantified Logics, is equivalent to that system.*

***Keywords**: Solving Necessary Equivalences, Modal Logic, Nonmonotonic Logic.*

1. Introduction

Solving equations is an important aspect of Automated Deduction because the solutions to an equation can often be used in other parts of a theorem in order to prove the theorem or to derive consequences. Normally, we think of solving equations in some numeric valued algebra such as number theory, real algebra, complex algebra, or linear spaces; but there is no reason the process cannot be applied to radically different types of mathematical structures although the actual techniques for solving equations will depend on the nature of that structure's algebra. A mathematical structure of particular interest in Artificial Intelligence is the sets of sentences of First Order Logic (i.e. FOL) which are deductively complete by the laws of FOL because such sets form the basis of the fixed-point theories of nonmonotonic reasoning such as Autoepistemic Logic [Moore 1985], its kernel [Konolige 1987], Reflective Logic [Brown 1987], and Default Logic [Reiter 1980]. A set of sentences of FOL is said to be deductively complete if and only if all the theorems deducible from it by the laws of FOL are contained within it. Deriving properties of infinite sets of sentences would appear to involve a sophisticated automatic theorem prover for set theory and FOL syntax. However, we avoid this by noting that deductively complete sets of sentences may be represented by the proposition which is the meaning of all the sentences in that set. The modal sentence $[]((\wedge_i\Gamma_i)\rightarrow\alpha)$ where [] is the necessity symbol can then be used to represent the proof theoretic statement $`\alpha\in(\mathrm{fol}\{`\Gamma_i\})$ where $`\alpha$ is the name of a sentence $\alpha$, for each i, $`\Gamma_i$ is the name of the sentence Γ_i of FOL, and fol is the set of FOL sentences derivable from $\{`\Gamma_i\}$ by the laws of FOL. A nonmonotonic fixed-point equation defined in set theory such as:

$$K=\mathrm{fol}(\{`\Gamma_i\}\cup\{`\alpha: (`(\neg\alpha)\notin K))\}$$

is then represented as the necessary equivalence:

$$k\equiv(\Gamma\wedge((\neg[]\neg(k\wedge\alpha))\rightarrow\alpha))$$

This algebra for solving for k in a necessary equivalence is just FOL supplemented with propositional variables (e.g. k above) and the necessity operator of a particular modal quantificational logic called Z. Translations of Autoepistemic Logic, Reflective Logic, and Default Logic, to Z are proven, respectively, in [Brown 2003c], [Brown 2003b], and [Brown 2003a]. An Automatic Deduction System, based on Z, for Autoepistemic Logic and Reflective Logic is given in [Brown 2003d]. Besides providing an algebra for representing such nonmonotonic systems (where quantified variables are not allowed to cross modal scopes), this modal representation has the additional advantage of providing an explication of what it means to have quantified variables crossing modal scopes:

$$k\equiv(\Gamma\wedge\forall\xi((\neg[]\neg (k\wedge\alpha))\rightarrow\alpha))$$

where ξ may occur free in α. Quantified Autoepistemic Logic, its Quantified Kernel, and Quantified Reflective Logic and their relationships are discussed in [Brown 2004]. Quantified Reflective Logic and Quantified Default Logic and their relationship is discussed in [Brown 2003e]. Herein, we exemplify necessary equivalence solving by solving an example involving normal defaults, similar to the example given above, which is expressible in all of these quantified systems.

Section 2 describes the Z Modal Quantificational Logic which is the algebra in which the necessary equivalence solving takes place. Section 3 discusses proving what is logically possible. Section 4 discusses necessary equivalence solving. Section 5 discusses deducing what is common to all solutions. Finally, some conclusions are drawn in Section 6.

2. The Modal Quantificational Logic Z

The syntax of Z Modal Quantificational Logic is an amalgamation of 3 parts:

(1) The first part is a First Order Logic (i.e. FOL) represented as the six tuple: ($\rightarrow$, #f, $\forall$, *vars*, *predicates*, *functions*) where $\rightarrow$, #f, $\forall$, are logical symbols, *vars* is a set of object variable symbols, *predicates* is a set of predicate symbols each of which has an implicit arity specifying the number of terms associated with it, and *functions* is a set of function symbols each of which has an implicit arity specifying the number of terms associated with it. Roman letters x, y, and z, possibly indexed with digits, are also variables.

(2) The second part is the extension to allow Propositional Quantifiers which consists of a set of propositional variables *propvars* and quantification over propositional variables (using $\forall$). Roman letters (other than x, y and z) possibly indexed with digits are used as propositional variables.

(3) The third part is Modal Logic [Lewis 1918] which adds the necessity symbol: [].

Greek letters are used as syntactic metavariables. $\pi,\pi_1...\pi_n,\rho,\rho_1...\rho_n$ range over the predicate symbols, ϕ, $\phi_1...\phi_n$ range over function symbols, $\delta,\delta_1...\delta_n$ range over terms, $\gamma,\gamma_1,...\gamma_n$, range over the object variables ξ, $\xi_1\ldots\xi_n,\zeta,\zeta_1\ldots\zeta_n$ range over a sequence of object variables of an appropriate arity, $f,f_1\ldots f_n$ range over predicate variables, and $\alpha,\alpha_1...\alpha_n$, $\beta,\beta_1...\beta_n$, $\chi,\chi_1...\chi_n$, Γ, κ range over sentences. Thus, terms are of the forms: γ and (ϕ $\delta_1...\delta_n$), and sentences are of the forms: ($\alpha \rightarrow\beta$), #f, ($\forall\gamma$ α), (π $\delta_1...\delta_n$), (f $\delta_1...\delta_n$), ($\forall f$ α), and ([]α). A zero arity predicate π, propositional variable f, or function ϕ is written as a sentence or term without parentheses, i.e., π instead of (π), f instead of (f), and ϕ instead of (ϕ). $\alpha\{\pi/\lambda\xi\beta\}$ is the sentence obtained from α by replacing all unmodalized occurrences of π by $\lambda\xi\beta$ followed by lambda conversion. $\alpha\{\pi_i/\lambda\xi\beta_i\}_{i=1,n}$, abbreviated as $\alpha\{\pi_i/\lambda\xi\beta_i\}$, represents simultaneous substitutions. $\wedge_{i=1,n}\beta_i$ represents ($\beta_1\wedge\ldots\wedge\beta_n$). The primitive symbols are listed in Figure 1. The defined symbols are listed in Figure 2.

Symbol	Meaning
$\alpha\rightarrow\beta$	if α then β.
#f	falsity
$\forall\gamma$ α	for all γ, α.
$\forall f$ α	for all f, α.
[] α	α is logically necessary

Figure 1: Primitive Symbols

Symbol	Definition	Meaning
$\neg\alpha$	$\alpha\rightarrow$ #f	not α
#t	$\neg$ #f	truth
$\alpha\vee\beta$	$(\neg\alpha)\rightarrow\beta$	α or β
$\alpha\wedge\beta$	$\neg(\alpha\rightarrow\neg\beta)$	α and β
$\alpha\leftrightarrow\beta$	$(\alpha\rightarrow\beta)\wedge(\beta\rightarrow\alpha)$	α if and only if β
$\exists\gamma$ α	$\neg\forall\gamma\neg\alpha$	some γ is α
$\exists f\alpha$	$\neg\forall f\neg\alpha$	some f is α

$<> \alpha$	$\neg[]\neg\alpha$	α is logically possible
$[\beta]\ \alpha$	$[](\beta\rightarrow\alpha)$	β entails α
$<\beta>\ \alpha$	$<>(\beta\wedge\alpha)$	α is possible with β
$\alpha\equiv\beta$	$[](\alpha\leftrightarrow\beta)$	α and β are synonymous
$\delta_1=\delta_2$	$(\pi\ \delta_1)\equiv(\pi\ \delta_2)$	δ_1 necessarily equals δ_2
$\delta_1\neq\delta_2$	$\neg(\delta_1=\delta_2)$	δ_1 does not necessarily equal δ_2
$(\text{det}\ \alpha)$	$\forall f(([\alpha]f)\vee([\alpha]\neg f))$where f is not in α	α is deterministic
$(\text{world}\ \alpha)$	$(<>\alpha)\wedge(\text{det}\ \alpha)$	α is a world

Figure 2: Defined Symbols of Z

The laws of Z Modal Logic is an amalgamation of five parts:

(1) The first part, given in Figure 3, consists of the laws of a FOL [Mendelson 1964].

(2) The second part, which is given in Figure 4, consists of the additional laws needed for propositional quantification. The laws SOLR2, SOLA4 and SOLA5 are the analogues of FOLR2, FOLA4 and FOLA5 for propositional variables.

(3) The Third part, which is given in Figure 5, consists of the laws MR0, MA1, MA2 and MA3 which constitute an S5 modal logic [Hughes & Cresswell 1968] [Carnap 1956]. When added to parts 1 and 2, they form a fragment of a Second Order Modal Quantificational logic similar to a second order version of [Bressan 1972].

(4) The fourth part, which is given in Figure 6, consists of the Priorian World extension of S5 Modal Logic. The PRIOR law [Prior and Fine 1977] states that a proposition is logically true if it is entailed by every world. This law was implied in [Leibniz 1686]

(5) The fifth part, which is given in Figure 7, consists of laws axiomatizing what is logically possible. MA1a lets one derive theorems such as $<>\forall x((\pi\ x)\leftrightarrow(x=\phi))$ and therefore extends the work on propositional possibilities in [Hendry and Pokriefka 1985] to possibilities in FOL being simpler than the FOL approaches in [Brown 1987]. MA1b and MA4 allow one to derive $(\phi_1\ \xi_1)\neq(\phi_2\ \xi_2)$ when ϕ_1 and ϕ_2 are distinct function symbols. MA4 states that at least two things are not necessarily equal just as there are at least two propositions: #t and #f.

FOLA3: $\alpha\rightarrow(\beta\rightarrow\alpha)$

FOLA4: $(\alpha\rightarrow(\beta\rightarrow\rho))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\rho))$

FOLA5: $((\neg\alpha)\rightarrow(\neg\beta))\rightarrow(((\neg\alpha)\rightarrow\beta)\rightarrow\alpha)$

FOLA6: $(\forall\gamma\ \alpha)\rightarrow\beta$ where β is the result of substituting an expression (which is free for the free positions of γ in α) for all the free occurrences of γ in α.

FOLA7: $(\forall\gamma\ (\alpha\rightarrow\beta))\rightarrow(\alpha\rightarrow(\forall\gamma\ \beta))$ where γ does not occur in α.

FOLR1: from α and $(\alpha\rightarrow\beta)$ infer β

FOLR2: from α infer $(\forall\gamma\ \alpha)$

Figure 3: The Laws of FOL

SOLR2: from α infer $(\forall f\ \alpha)$

SOLA6: $(\forall f\ \alpha)\rightarrow\beta$ where β is the result of substituting an expression (which is free for the free positions of f in α) for all the free occurrences of f in α.

SOLA7: $(\forall f(\alpha\rightarrow\beta))\rightarrow(\alpha\rightarrow(\forall f\ \beta))$ where f does not occur in α.

Figure 4: Additional Laws for Propositional Quantifiers

MA1: $([]\alpha)\rightarrow\alpha$

MA2: $([\alpha]\beta)\rightarrow(([]\alpha)\rightarrow([]\beta))$

MR0: from α infer $([]\ \alpha)$ provided α was derived only from logical laws

MA3: $([]\alpha)\vee([]\neg[]\alpha)$

Figure 5: Additional Laws of S5 Modal logic

PRIOR: $(\forall w((\text{world}\ w)\rightarrow([w]\alpha)))\rightarrow[]\alpha$ where w does not occur free in α.

Figure 6: Additional Law of a Priorian World Logic

MA1a: $([]\alpha)\rightarrow\alpha\{\pi/\lambda\zeta\beta\}$ if no unmodalized occurrence of a higher order var is in α.

MA1b: $([]\alpha) \rightarrow \alpha\{\phi/\lambda\zeta\delta\}$ where α may not contain an unmodalized occurrence of a higher order variable, nor an unmodalized free object variable.

MA4: $\exists x\exists y(x\neq y)$

Figure 7: Additional Laws of Z Modal Logic

3. Deriving what is Logically Possible

The main problem with using axiom scheme MA1a to prove that something is logically possible lies in finding the appropriate substitution for the parameter π. The theorem ZP1 given below is the basis of a heuristic for finding such instances. Intuitively, we know that the conjunction of instances of a predicate and the conjunction of instances of the negation of a predicate are logically possible whenever the two do not coincide on any instance. For example: where a≠c and b≠c then $((\pi\ a)\wedge(\pi\ b)\wedge(\neg(\pi\ c)))$ is logically possible but $((\pi\ a)\wedge(\pi\ b)\wedge(\neg(\pi\ a)))$ is not. Thus if a sentence can be written in the form: $(\Gamma\wedge(\forall x(\alpha\rightarrow(\pi\ x)))\wedge(\forall x(\beta\rightarrow(\neg(\pi\ x)))))$ where π does not occur in α, β, and Γ, then it is logically possible if and only if $\{x{:}\alpha\}\cap\{x{:}\beta\}$ is empty, which is to say that $\exists x(\alpha\wedge\beta)$ does not follow from Γ or to say that Γ and $\neg\exists x(\alpha\wedge\beta)$ is logically possible. In this manner determining whether a sentence with n predicates is logically possible can sometimes be reduced to determining whether a sentence with n-1 predicates is logically possible without having to guess any instances of π in MA1a.

Theorem ZP1: The Possibility of a Disjoint Predicate Definition: [Brown 1989]

If Γ, α, and β do not contain any unmodalized occurrences of π nor of any higher order variable then:

$(<>(\Gamma\wedge(\forall\xi(\alpha\rightarrow(\pi\ \xi)))\wedge(\forall\xi(\beta\rightarrow\neg(\pi\ \xi))))) \leftrightarrow(<>(\Gamma\wedge(\neg\exists\xi(\alpha\wedge\beta))))$

ZP1 is applicable to any theory which can be put into a prenix conjunctive normal form such that no disjunct contains more than one unmodalized occurrence of π, since by the laws of classical logic such theories are equivalent to an expression of the form: $(\Gamma\wedge(\forall\xi(\alpha\rightarrow(\pi\ \xi)))\ \wedge(\forall\xi(\beta\rightarrow\neg(\pi\ \xi))))$. It is also applicable to any theory which can be put into a disjunction of a prenix conjunctive normal forms whose disjunctions contains no more than one unmodalized occurrence of π. The reason for this is that disjunction (and existential quantifiers) associate through possibility: $(<>(\alpha\vee\beta))\leftrightarrow((<>\alpha)\vee(<>\beta))$. For this reason, ZP1 is decidable for many important cases of FOL including propositional logic and the case of a finite theory.

Example 1: Deducing a Logical Possibility thrice using ZP1. Let k be any sentence.

$<>((P\ b)\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$
$\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

This is equivalent to:

$<>((P\ b)\wedge(P\ a)\wedge\forall x\ ((x=b)\rightarrow\neg(Q\ x))\wedge\forall x(((\neg(AB\ x))\wedge(P\ x))\rightarrow(Q\ x))$
$\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

Instantiating ZP1 by letting α be $\forall x(((\neg(AB\ x))\wedge(P\ x))\rightarrow(Q\ x))$, β be $\forall x\ ((x=b)\rightarrow\neg(Q\ x))$ and Γ be the other sentences qives:

$<>((P\ b)\wedge(P\ a)\wedge\neg\exists x((\neg(AB\ x))\wedge(P\ x)\wedge(x=b))\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

which is equivalent to: $<>((P\ b)\wedge(P\ a)\wedge(AB\ b))\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

which is: $<>((P\ b)\wedge(P\ a)\wedge\forall x((x=b)\rightarrow(AB\ x))\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

Instantiating ZP1 by letting α be $\forall x((x=b)\rightarrow\neg(Q\ x))$, β be $\forall x((x\neq b)\rightarrow\neg(AB\ x))$, and Γ be the other sentences qives: $<>((P\ b)\wedge(P\ a)\wedge\neg\exists x((x=b)\wedge(x\neq b))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

which is equivalent to: $<>((P\ b)\wedge(P\ a)\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$

which is: $<>(\#t\wedge\forall x(((x=a)\vee(x=b)\vee((x\neq a)\wedge(x\neq b)\wedge([k](P\ x))))\rightarrow(P\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b)\wedge(\neg([k](P\ x))))\rightarrow\neg(P\ x))$

Instantiating ZP1 by letting α be $((x=a)\vee(x=b)\vee((x\neq a)\wedge(x\neq b)\wedge([k](P\ x))))$, β be $((x\neq a)\wedge(x\neq b)\wedge(\neg([k](P\ x))))$ and Γ be #t gives: $\langle\rangle(\#t\wedge\neg\exists x(((x=a)\vee(x=b)\vee((x\neq a)\wedge(x\neq b)\wedge([k](P\ x))))\wedge((x\neq a)\wedge(x\neq b)\wedge(\neg([k](P\ x))))\ ))$ which $\langle\rangle\#t$ which is #t.

4. Solving Necessary Equivalences when Variables Cross Modal Scopes

A necessary equivalence has the form $k\equiv(\Phi\ k)$ where k is a propositional variable. The goal is to transform the initial necessary equivalence into a (possibly infinite) disjunction, $\exists i(k\equiv\beta_i)$ with each β_i free of k. If $k\equiv\beta_i$ implies the original equation then β_i is a solution to the original equation. The following procedure solves necessary equivalences when no variable crosses modal scopes:

Procedure for Solving Modal Equivalences [Brown 1986]:

Step 1: First, one by one, each subformula α which contains k and is equivalent to $([]\alpha)$ is pulled out of the necessary equivalence causing it to be split into two cases. This is done by using the following theorem schema replacing any instance of the left side by the corresponding instance of the right side:

$(k\equiv(\phi\ \alpha))\leftrightarrow((\alpha\wedge(k\equiv(\phi\ \#t)))\vee((\neg\alpha)\wedge(k\equiv(\phi\ \#f))))$

Step 2: Second, the resulting equivalences are simplified by the laws of the modal logic.

Step 3: Third, on each disjunct the simplified value for k is back substituted into each such α or $(\neg\alpha)$ sentence thereby eliminating k from them.

Step 4: Fourth, the α and $(\neg\alpha)$ sentences are simplified using the modal laws giving a disjunction of necessary equivalences..

When no variables cross modal scopes, for any decidable case of First Order Logic, this method is an algorithm resulting in a finite disjunction of solutions as is illustrated in Example 2 below:

Example 2: Solving a Modal Equation: $k\equiv(((\langle k\rangle\neg A)\rightarrow B)\wedge((\neg\langle k\rangle\neg A)\rightarrow A))$

Step 1 gives: $((\langle k\rangle\neg A)\wedge(k\equiv((\#t\rightarrow B)\wedge((\neg\#t)\rightarrow A))))\vee((\neg(\langle k\rangle\neg A))\wedge(k\equiv((\#f\rightarrow B)\wedge((\neg\#f)\rightarrow A))))$

Step 2 gives: $((\langle k\rangle\neg A)\wedge(k\equiv B))\vee((\neg(\langle k\rangle\neg A))\wedge(k\equiv A))$

Step 3 gives: $((\langle B\rangle\neg A)\wedge(k\equiv B))\vee((\neg(\langle A\rangle\neg A))\wedge(k\equiv A))$

Step 4 gives: $(\#t\wedge(k\equiv B))\vee(\#t\wedge(k\equiv A))$ which is: $(k\equiv B)\vee(k\equiv A)$

The process described above provides an algorithm [Brown & Araya 1991] for solving necessary equivalences in the modal representations of Reflective Logic [Brown 2003a] and Autoepistemic Logic [Brown 2003c], and with some additional details of [Default Logic 2003b]. These modal representations can be generalized to allow for universally quantified variables crossing modal scopes (at the top level of the right side of the equation. We now address the problem of solving necessary equivalences when quantified variables cross modal scopes as in:

$$k\equiv(\Gamma\wedge\wedge_{\square}\forall\xi_{\square}((([k]\alpha_{\square})\wedge(\wedge_{\square=1,mi}(\langle k\rangle\beta_{\square\square})))\rightarrow\chi_{\square})$$

We want to eliminate the modal expressions containing k from the right side of the necessary equivalence. The difficulty lies in the fact that in general we cannot apply Step 1 in the above algorithm because the modal expressions may contain the $\xi_{\square}$ variables which are captured by quantifiers inside the right side of the necessary equivalence. The solution to this dilemma is to allow quantified statements such as: $\forall\xi((([k]\alpha)\wedge(\wedge_{j=1,m}(\langle k\rangle\beta_j)))\rightarrow\chi)$ to be divided into a finite number of instances for each particular formula, which may or may not hold in a particular solution, leaving all the remaining instances in the quantified statement:

Step 0: Divide a quantified expression over a modal scope into parts using the schema:

$$(k\equiv(\Psi\wedge\forall\xi((([k]\alpha)\wedge(\wedge_{j=1,m}(\langle k\rangle\beta_j)))\rightarrow\chi_{\square})))$$

$$\leftrightarrow(k\equiv(\Psi\wedge\forall\xi((\phi\wedge([k]\alpha)\wedge(\wedge_{j=1,m}(\langle k\rangle\beta_j)))\rightarrow\chi_{\square}))\wedge\forall\xi((\neg\phi\wedge([k]\alpha)\wedge(\wedge_{j=1,m}(\langle k\rangle\beta_j)))\rightarrow\chi)))$$

where ϕ specifies a finite number of instances thereby allowing the quantifier above it in the resulting expression to be eliminated. The parts remaining under the quantifier that are consistent with the solution may then sometimes be eliminated by theorems PR or NPR. We call this application Step 5:

Theorem PR: Reduction of Possible Reflections: If Γ, α, and β are sentences of Z, and if γ is not free in Γ then:

(∀γ <>(Γ∧(∀γ β)∧α)) →((k≡(Γ∧∀γ((<k>α)→β)))↔(k≡(Γ∧∀γ β)))

proof: (k≡(Γ∧∀γ ((<k>α)→β)))↔ (k≡(Γ∧∀γ β)) is true iff:

((k≡(Γ∧∀γ ((<k>α)→β))) → (k≡(Γ∧∀γ β))) ∧ ((k≡(Γ∧∀γ β)) → (k≡(Γ∧∀γ ((<k>α)→β))))

Using the hypothesis in each case we get:

(((k≡(Γ∧∀γ ((<k>α)→β)))→((Γ∧∀γ((<k>α)→β))≡(Γ∧∀γ β)))

∧((k≡(Γ∧∀γ β))→ ((Γ∧∀γ β)≡(Γ∧∀γ ((<k>α)→β)))))

which would be true if: (((k≡(Γ∧∀γ ((<k>α)→β)))→∀γ(<k>α)) ∧ ((k≡(Γ∧∀γ β))→∀γ (<k>α)))

which is implied by: (∀γ (<(Γ∧∀γ((<k>α)→β))>α))∧∀γ(<(Γ∧∀γ β)>α)

which is equivalent to just: ∀γ (<>(Γ∧(∀γ β)∧α)) which is the hypothesis of the theorem. QED.

Theorem NPR: Reduction of Normal Possible Reflections:

If Γ and α are sentences of Z, and if γ is not free in Γ then:

(∀γ <>(Γ∧(∀γ α))) →((k≡(Γ∧∀γ ((<k>α)→α)))↔(k≡(Γ∧∀γ α)))

proof: Letting β be α in PR gives: (∀γ <>(Γ∧(∀γ α)∧α)) → ((k≡(Γ∧∀γ((<k>α)→α)))↔(k≡(Γ∧∀γ α)))

Since (∀γ α) implies α this is proves the theorem. QED.

Example 3: Partially Solving a Modal Equation with Quantifiers over Modal Scopes

k≡((P a)∧(¬(Q b))∧∀x((¬(AB x))→((P x)→(Q x))) ∧∀x((<k>¬(AB x))→¬(AB x)) ∧∀x((P x)↔([k](P x))))

Step 0: Dividing ∀x((P x)↔([k](P x))) on a and b and simplifying by noting that [k]Pa gives :

(k≡((P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x))) ∧∀x((<k>¬(AB x))→¬(AB x))

∧((P b)↔([k](P b)))∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x))))))

Steps 1&2: Splitting on ([k](P b)) gives:

((([k](P b))∧(k≡((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x))) ∧∀x((<k>¬(AB x))→¬(AB x))

∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x)))))))

∨(¬([k](P b))∧(k≡((¬(P b))∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))

∧∀x((<k>¬(AB x))→¬(AB x)) ∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x))))))))

Step 0: In the first necessary equivalence (P b)∧(¬(Q b)) implies (AB b). Thus dividing the AB default ∀x((<k>¬(AB x))→¬(AB x)) on b gives: ∀x(((x≠b)∧(<k>¬(AB x)))→¬(AB x)) which is equivalent to: (<k>((x≠b)→¬(AB x)))→((x≠b)→¬(AB x)) resulting in:

((([k](P b))∧(k≡((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))

∧∀x((<k>((x≠b)→¬(AB x)))→((x≠b)→¬(AB x)))∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x)))))))

∨(¬([k](P b))∧(k≡((¬(P b))∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))

∧∀x((<k>¬(AB x))→¬(AB x)) ∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x))))))))

In the second necessary equivalence, if a=b then k≡#f which implies that ¬([k](P b)) is equivalent to ¬([#f](P b)) which is equivalent to #f. Thus a≠b on the second case giving:

((([k](P b))∧(k≡((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))

∧∀x((<k>((x≠b)→¬(AB x)))→((x≠b)→¬(AB x))∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x)))))))

∨(¬([k](P b))∧(a≠b)∧(k≡((¬(P b))∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))

∧∀x((<k>¬(AB x))→¬(AB x)) ∧∀x(((x≠a)∧(x≠b))→((P x)↔([k](P x))))))))

Step 5 twice: NPR is used to eliminate (<k>((x≠b)→¬(AB x))) from the first necessary equivalence and (<k>¬(AB x)) from the second necessary equivalence. On the first necessary equivalence the hypothesis to NPR is: <>((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x((x≠b)→¬(AB x))∧∀x(((x≠a)∧(x≠b)) →((P

$x)\leftrightarrow([k](P\ x)))))$ and on the second necessary equivalence the hypothesis to NPR is: $<>((\neg(P\ b))\wedge(P\ a)\wedge(\neg(Q\ b))\wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))\wedge\forall x(\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )$. The first possibility is deduced to be #t by three applications of ZP1 (See Example 1 herein). The second possibility is deduced to be $a\neq b$ by three applications of ZP1. Since $a\neq b$ is a hypothesis of this case it is true. Applying NPR in both cases then gives:

$$(((([k](P\ b))\wedge(k\equiv((P\ b)\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$$
$$\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )))$$
$$\vee(\neg([k](P\ b))\wedge(a\neq b)\wedge(k\equiv((\neg(P\ b))\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$$
$$\wedge\forall x(\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ ))))$$

Steps 3&4 twice: Since (P b) is in the first necessary equivalence, the entailment on the first case holds. Likewise since $(\neg(P\ b))$ is in the second necessary equivalence $(\neg([k](P\ b)))$ is $(\neg([k]\#f))$ which is $(<>k)$ which holds since #f is not a solution if $a\neq b$. Thus we get:

$$(((k\equiv((P\ b)\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$$
$$\wedge\forall x((x\neq b)\rightarrow\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ )))$$
$$\vee((a\neq b)\wedge(k\equiv((\neg(P\ b))\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$$
$$\wedge\forall x(\neg(AB\ x))\wedge\forall x(((x\neq a)\wedge(x\neq b))\rightarrow((P\ x)\leftrightarrow([k](P\ x))))\ ))))$$

Since (AB b) is derivable in the first necessary equivalence, and since in either equation (P a) and (P b) hold if and only if each is entailed in k this is equivalent to:

$$((k\equiv((P\ b)\wedge(P\ a)\wedge(\neg(Q\ b))\wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))\wedge\forall x((AB\ x)\leftrightarrow(x=b))\wedge\forall x((P\ x)\leftrightarrow([k](P\ x)))\ ))$$
$$\vee((a\neq b)\wedge(k\equiv((\neg(P\ b))\wedge(P\ a)\wedge(\neg(Q\ b))\ \wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))\wedge\forall x(\neg(AB\ x))\ \wedge\forall x((P\ x)\leftrightarrow([k](P\ x)))))))$$

The result is essentially a disjunction of two necessary equivalences (which partially solve for k. Further derivation along these lines gives solutions making P hold for any subset of {x: #t}-{a b}. However, if all we want is what is common to all solutions then eliminating ([k](P x)) is not required as is shown in the next section.

5. Aggregating the Solutions

Parallel Circumscription [McCarthy 1986] in Second Order Logic with circumscribed π_i, variable, and fixed predicates ρ_j is equivalent to the "infinite disjunction" of all the solutions of a necessary equivalence:

$$\exists k(k\wedge(k\equiv(\Gamma\wedge\ \wedge_i\forall\xi_i((<k>\neg(\pi_i\ x))\rightarrow\neg(\pi_i\ x))\ \wedge\ \wedge_j\forall\xi_j((\rho_j\ x)\leftrightarrow([k](\rho_j\ x)))\ )))$$

For this reason, in some cases, such as in Example 3 above we can compare the disjunction of the solutions with results obtained by Circumscription. The key theorem for this comparison is the following theorem which allows fixed predicates to be ignored after the Circumscribed predicates are eliminated.

<u>The Fixed Predicate Lemma [Brown 1989]:</u> $\exists k(k\wedge(k\equiv(\Gamma\wedge\ \wedge_i\forall\xi((\rho_i\ x)\leftrightarrow([k](\rho_i\ x))))))\equiv\Gamma$

The necessary equivalences related to Circumscription are a small subclass of the necessary equivalences that are expressible. The main goal herein is to solve necessary equivalences, rather than to compute the "infinite disjunction" of solutions which is all that Circumscription does. However, the solutions in Example 3 are aggregated and compared in Example 4 with Circumscription for the case where $a\neq b$. This case encompasses the case where a and b are distinct 0-arity function symbols.[10]

<u>Example 4: Aggregating the Solutions when $a\neq b$ to $(P\ a)\wedge(\neg(Q\ b))\wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$ by AB with P fixed and Q variable:</u>

$$\exists k(k\wedge(k\equiv((P\ a)\wedge(\neg(Q\ b))\wedge\forall x((\neg(AB\ x))\rightarrow((P\ x)\rightarrow(Q\ x)))$$
$$\wedge\forall x((<k>\neg(AB\ x))\rightarrow\neg(AB\ x))\wedge\forall x((P\ x)\leftrightarrow([k](P\ x)))\ )))$$

From Example 3 and assuming $a\neq b$ we get:

[10] Herein, = is necessary equality as defined in Figure 2. It should not be confused with an extensional equality predicate which only provides substitution properties through nonmodal contexts.

∃k(k∧((k≡((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x((AB x)↔(x=b))∧∀x((P x)↔([k](P x)))))

∨(k≡((¬(P b))∧(P a)∧(¬(Q b))∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x(¬(AB x)) ∧∀x((P x)↔([k](P x)))))))

Pushing ∃k to lowest scope gives:

((∃k(k∧(k≡((P b)∧(P a)∧(¬(Q b))∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x((AB x)↔(x=b))∧∀x((P x)↔([k](P x)))))))

∨(∃k(k∧(k≡((¬(P b))∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x(¬(AB x))∧∀x((P x)↔([k](P x))))))))

By the Fixed Predicate Lemma (twice) this is equivalent to:

((P b)∧(P a)∧(¬(Q b)) ∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x((AB x)↔(x=b)))

∨((¬(P b))∧(P a)∧(¬(Q b))∧∀x((¬(AB x))→((P x)→(Q x)))∧∀x(¬(AB x)))

which is equivalent to:

(P a)∧(¬(Q b))∧∀x((¬(AB x))→((P x)→(Q x)))∧(((P b)∧∀x((AB x)↔(x=b)))∨((¬(P b))∧∀x(¬(AB x))))

Since a≠b, It follows that ¬(AB a) and therefore that (Q a) holds. This is exactly what one would expect as is suggested by the following quote from [Konolige 1989]: "Consider the simple abnormality theory (see [McCarthy 1986]), with W ={∀x.Px ∧¬AB(x) → Qx, Pa, ¬Qb} (this is a variation of an example in [Perlis, 1986].) We would expect Qa to be a consequence of Circum(W;ab;Q), but it is not."[11] "The reason is that there are ab-minimal models of W in which b and a refer to the same individual and ¬Qa is true."[12]

6. Conclusion

The Z Modal Quantificational Logic provides an interesting algebra for deriving fixed-point solutions to necessary equivalences where universally quantified variables cross modal scope. Herein, some specific methods for solving some simple classes of problems have been described and exemplified. The presented methods do not solve all problems, but the Z logic provides a framework for developing more general solution methods generalizing the ones herein presented. Since many Quantified Nonmonotonic Logics are representable in the Z Modal Logic, including Quantified Autoepistemic Logic, Quantified Autoepistemic Kernels, Quantified Reflective Logic, and Quantified Default Logic, such deduction techniques could be applicable to a wide range of Quantified generalizations of most of the well known nonmonotonic logics.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Numbers: MIP-9526532, EIA-9818341, and EIA-9972843.

Bibliography

Bressan, Aldo 1972. *A General Interpreted Modal Calculus,* Yale University Press.

Brown, F. M. 1986. "A Commonsense Theory of NonMonotonic Reasoning", *Proceedings of the 8h International Conference on Automated Deduction, CADE8,* Lecture Notes in Computer Science, vol 230, Oxford, England, July 1986, Springer Verlag, New York.

Brown, Frank M. 1987. "The Modal Logic Z", In *The Frame Problem in AI"; Proceedings of the 1987 AAAI Workshop*, Morgan Kaufmann, Los Altos, CA .

Brown, Frank M. 1989. "The Modal Quantificational Logic Z Applied to the Frame Problem", advanced paper *First International Workshop on Human & Machine Cognition*, May 1989 Pensacola, Florida. Abbreviated version published in *International Journal of Expert Systems Research and Applications, Special Issue: The Frame Problem. Part A.* eds.

[11] If a≠b, such as is the case where a and b are two distinct 0-arity function symbols, it is a consequence of Circumscription as defined in Second Order Logic. In Second Order Logic x=y is defined to be ∀ƒ((ƒ x)↔(ƒ y)).

[12] The obvious definition of minimal model does not give the desired properties. The definition of minimal model could be changed to being minimal not with respect to all models but with respect to those models giving function symbols the same interpretation.

Kenneth Ford and Patrick Hayes, vol. 3 number 3, pp169-206, JAI Press 1990. Reprinted in *Reasoning Agents in a Dynamic World: The Frame problem*, editors: Kenneth M. Ford, Patrick J. Hayes, JAI Press 1991.

Brown, Frank M.. 2003a "Representing Reflective Logic in Modal Logic", *Information Theories and Applications*, Vol. 10, no. 4, pages: 431-438, December 2003.

Brown, Frank M.. 2003b, "Representing Default Logic in Modal Logic", *Information Theories and Applications*, Vol. 10, no. 4, pages 439-446, December, 2003.

Brown, Frank M.. 2003c, "Representing Autoepistemic Logic in Modal Logic", *Information Theories and Applications*, Vol. 10, no. 4, pages 455-462, December 2003.

Brown, Frank M.. 2003d, "On the Relationship between Quantified Reflective Logic and Quantified Default Logic" *Information Theories and Applications*, Vol. 10, no. 4, pages 447-454, December 2003.

Brown, Frank M. 2003e, "Decision Procedures for the Propositional Cases of 2nd Order Logic and Z Modal Logic Representations of a 1st Order L-Predicate Nonmonotonic Logic", *Automated Reasoning with Analytic Tableaux and Related Method: Tableaux 2003*, September 2003, Rome, *Lecture Notes in Artificial Intelligence 2796*, ISBN 3-540-40787-1, Springer-Verlag, Berlin, 2003.

Brown, Frank, 2004, "On the Relationships Among Quantified Autoepistemic Logic, its Kernel, and Quantified Reflective Logic", to appear in Int. Conf. i.TECH 2004.

Brown,Frank M. and Carlos Araya 1991. "A Deductive System for Theories of NonMonotonic Reasoning," *Transactions of the Eighth Army Conference on Applied Mathematics and Computing, 1991*.

Carnap, Rudolf 1956. *Meaning and Necessity: A Study in the Semantics of Modal Logic*, The University of Chicago Press.

Hendry, Herbert E. and Pokriefka, M. L. 1985. "Carnapian Extensions of S5", *Journal of Phil. Logic 14*.

Hughes, G. E. and Cresswell, M. J., 1968. *An Introduction to Modal Logic*, Methuen & Co. Ltd., London.

Konolige, Kurt 1989. "On the Relation between Autoepistemic Logic and Circumscription Preliminary Report", IJCAI89.

Leibniz, G. 1686. "Necessary and Contingent Truths", *Leibniz Philosophical Writings*, Dent&Sons. 1973.

Lewis, C. I. 1936. Strict Implication, *Journal of Symbolic Logic*, vol I..

McCarthy, J. 1986, "Applications of Circumscription to Formalizing Common-Sense Reasoning", *Artificial Intelligence*, volume 28.

Mendelson, E. 1964. *Introduction to Mathematical Logic*, Van Norstrand, Reinhold Co., New York.

Moore, R. C. 1985. "Semantical Considerations on Nonmonotonic Logic" *Artificial Intelligence*, 25.

Prior, A. N. & Fine, k. 1977. *Worlds, Times, and Selves*, Gerald Duckworth.

Reiter, R. 1980. "A Logic for Default Reasoning" *Artificial Intelligence*, 13.

Author information

Frank M. Brown- University of Kansas, Lawrence, Kansas, 66045, e-mail: brown@ku.edu.

ABOUT NEW PATTERN RECOGNITION METHOD FOR THE UNIVERSAL PROGRAM SYSTEM "RECOGNITION"

A.A. Dokukin, O.V. Senko

***Abstract**: In this work the new pattern recognition method based on the unification of algebraic and statistical approaches is described. The main point of the method is the voting procedure upon the statistically weighted regularities, which are linear separators in two-dimensional projections of feature space. The report contains brief description of the theoretical foundations of the method, description of its software realization and the results of series of experiments proving its usefulness in practical tasks.*

***Keywords**: pattern recognition, statistically weighted regularities, voting procedure.*

Introduction

Nowadays there are a great number of effective pattern recognition methods based on voting procedure upon some kind of regularities in the data, as well as different approaches for searching these regularities. The term "regularity" is interpreted as some sub-region in space of prognostic variables where fraction of at least one of the classes differs significantly from its fraction in neighbor regions. For example, there are the method of voting upon the sets of irreducible tests [1] or representative tests [13], the method of voting upon statistically weighted syndromes (further in the text it is referred as SWS) [2], method of voting upon sets of logical regularities [3] and etc. Results of hands-on testing show the higher steadiness of voting procedure to the minor changes in training and testing samples, which leads to the significant increase of quality of voting-based methods. This advantage is especially important in relatively high-dimensional tasks with limited number of cases in data sets. The theoretical substantiation of this fact [4] exceeds the bounds of this report, but the detailed proof by the means of mathematical statistics is now being prepared for publication.

All these methods have one strong restriction and maybe even disadvantage. It is the fact that all regularities are some kind of hyper parallelepipeds in feature space with planes orthogonal to the datum lines. However in many tasks the essentially multidimensional regularities may arise which are separated from neighborhood by multivariate linear boundaries. So the method of two-dimensional linear separators (further referred as TLS) presents an attempt to complicate shape of elementary regularities preserving all advantages of voting procedure.

The Method of Two-Dimensional Linear Separators (TLS)

Further following notation will be used. Let's consider the set of permissible objects M , let's also consider that it presents Cartesian product of n sets of permissible values of features $M = M_1 \times ... \times M_n$. It is presumed that there is the unknown subdivision of the set M into l classes $K_1,...,K_l$. This subdivision is described by means of training sample $S_1,...,S_m$ of objects $S_i = a_{i1},...,a_{in}, i = 1,...,m$, for which the classification is known: $\alpha(S) = \alpha_1,...,\alpha_l$, where $\alpha_j =< S \in K_j >, j = \overline{1,l}$. It is necessary to restore the unknown classification of the testing sample $S^1,...,S^q$.

The main point of the method is the successive examination of different pairs of features and construction of the linear separator for every one of them and for every class. These separators must divide two-dimensional projections of objects of selected class and its additive inversion. For every class K_i and for every pair of features (u,v) the found line $L^i_{(u,v)}$ is called elementary regularity. Moreover the weight of the regularity $w^i_{(u,v)}$ is calculated due to the separating ability of the line.

The recognition is based on the weighted voting procedure by the set of elementary regularities. Let's consider the estimation for k-th class, the rest ones are calculated in much the same way. Each regularity $L^k_{(u,v)}$ refers the new object S to the k-th class or to its additive inversion, so the part of training objects of k-th class in the half plain there the object was referred to can be calculated for each pair of features $\upsilon^{k}_{(u,v)}(S)$. The final estimation is calculated according to the following formula:

$$\Gamma(S, K_1) = \frac{\sum_{(u,v)} \upsilon^{1}_{(u,v)}(S) w^{1}_{(u,v)}}{\sum_{(u,v)} w^{1}_{(u,v)}}.$$

The weights of regularities are calculated in much the same manner as in the method of statistically weighted syndromes (SWS) and depend on the quality of separating of training sample. If there are any errors in separation, i.e. some objects from the class are referred to its additive inversion or objects from inversion are referred to the class, the weight is set to be inversely proportional to the variance of the error

$$\frac{1}{p(1-p)},$$

there p is the part of errors. If this is not the case and all training objects are separated correctly than the variance of error is replaced with its Bayesian estimation

$$\frac{\int_0^1 (1-p)^n \, dp}{\int_0^1 (1-p)^n \, pdp}$$

In the TLS method the linear separators are sought by means of pattern recognition method called Linear Machine [5]. Its main point is that the task of finding separating line is replaced with the task of finding the maximal simultaneous subsystem of the system of linear inequalities and its subsequent solving by means of relaxation algorithm.

In conclusion of this paragraph let's consider the results of some hand-on testing. In the table 1 there are some tasks that clearly demonstrate the advantages of unification of voting procedure and complex elementary regularities. It contains the results of comparison of Linear Machine, SWS and TLS methods. The method with best performance is marked with gray color.

Task	LM	SWS	TLS
Breast	94.9	94.1	95.2
Ionosphere	85.2	90.1	90.1
Iris	97.5	95	97.5
Mel	50	65.6	68.8
Patomorphosis	76.5	85.3	91.2

Table 1 Comparison of LM, SWS ans TLS methods

The following tasks were considered during the test series:

- Breast – the breast cancer recognition, 9 features, 2 classes, 344 training examples, 355 testing ones (Breast cancer databases was obtained from Dr. William H. Wolberg from the University of Wisconsin Hospitals, Madison [6]);
- Ionosphere – the recognition of structural peculiarities in ionosphere, 34 features, 2 classes, 170 training examples, 181 testing ones (data from Johns Hopkins University Ionosphere database);
- Iris – Iris recognition, 4 features, 3 classes, 71 training examples, 81 testing ones (data from Michal Marshall's Iris Plants Database);
- Mel – Recognition of melanoma by the set of geometrical and radiological features, 33 features, 3 classes, 48 training examples, 32 testing ones [12];

- Patomorphosis – forecast of destruction level of malignant growth after chemotherapy by the set of parameters characterizing optical behavior of its cell nucleus, 7 features, 2 classes, 43 training examples and 31 testing ones (the data has been received from Dr. Matchak from Cancer Research Center of the Russian Academy of Medical Sciences).

In the table 2 the results of comparison of TLS with some other methods build-in to the Recognition software system are shown. The methods ate tested with two tasks which features are small number of objects in comparison with dimension of task. It is important that the suggested method has shown the significant increase of quality in this class of tasks.

Method	Mel	Patomorphosis
TLS	68.8	91.2
LM	50	76.5
SWS	65.6	85.3
LDF	59.4	76.5
AVO	62.5	76.5
IT	62.5	85.3
QNN	62.5	70.6
Perceptron	65.6	79.4
SVM	56.3	76.5

Table 2 Comparison with other methods on the tasks with short samples.

Following methods were used: TLS – Two-dimensional Linear Separators, LM – the mentioned above Linear Machine[5], SWS [2], LDF - Fisher's Linear Discriminant [7], AVO or ECA – Estimates Calculating Algorithm [8], IT – voting upon Irreducible tests [1], QNN – q Nearest Neighbors [7], Perceptron – Multilayer Perceptron [7,9], SVM – Support Vector Machine [10].

Software Realization

Software system "Recognition" has been developed in Dorodnicyn Computing Centre of Russian Academy of Sciences in cooperation with Solutions Ltd. The system's detailed description can be found, for example, in the proceedings of the Open German-Russian Workshop [11] or at the developer's Internet sight http://www.solutions-center.ru. In this article only the brief description of its basic principles will be given, because these principles have been being considered throughout the whole TLS' development process. They are universality, uniformity, modularity and intellectuality.

The universality of the system is understood as a wide coverage of different approaches to pattern recognition and classification including so-called classifier fusion, which are realized in the system's library of methods. The methods have been developed as separate interchangeable modules of uniform structure. On the software level each module is a single dynamic-link library with standardized interface.

While solving a wide variety of different practical task the initial assumption that each recognition method has its advantages and there is no single best one for every kind of tasks has been proven. So the main accent was made on using of different kinds of classifier fusions. And the uniformity of the methods allows combining the results of every subset of developed methods into one classifier, providing results more accurate than average and even close to maximal ones in automatic training mode. This fact allows claiming some kind of intellectuality of the developed system.

Passing on to software realization of Two-dimensional Linear Separators method itself, it is important to take note of two facts:

First of all, developing of TLS method's software realization was significantly simplified due to availability of software system "Recognition", since it was taking care of all the chores including preparation of methods environment, quality control and etc. So the developers in the person of the authors of this paper were able to get concentrated on the method itself.

Secondly, the fact that TLS method has significantly increased the quality of recognition for some kind of tasks has been already mentioned in section 2. Thus the addition of this method to classifier fusions allows increasing their quality greatly for these tasks. The experimental proof of this fact is shown in table 3. One of the simplest ways of constructing classifier fusion has been considered. The simple majority voting procedure has been applied firstly to the set of LM, SWS, LDF, AVO, IT, QNN, Perceptron and SVM, and secondly to the same set of algorithms in addition with TLS.

Task	Without TLS	With TLS
Mel	59.4	65.6
Patomorphosis	76.5	82.4

Table 3 Quality of majority voting

Conclusion

In conclusion we can claim that the developed method has justified our hopes. The combination of voting procedure and linear separators has increased recognition quality in some class of practically important tasks. Thus the developed software realization of TLS method can serve as a great support for researchers.

Acknowledgment

The research has been supported by RFBR grants NN 02-07-90137, 02-07-90134, 02-01-00558, 03-01-00580, 02-01-08007 and state program НШ-1721.2003.1 (scientific schools).

Also thanks goes to Dr. William H. Wolberg from the University of Wisconsin Hospitals, Madison; Vince Sigillito from Johns Hopkins University (vgs@aplcen.apl.jhu.edu); Michal Marshall (MARSHALL%PLU@io.arc.nasa.gov); Dr. Ganster from Linz University and Dr. Matchak from Cancer Research Center of Russian Academy of Medical Sciences for the granted data.

Bibliography

[1] А.Н.Дмитриев, Ю.И.Журавлев, Ф.П.Кренделев, О математических принципах классификации предметов и явлений. Сб. "Дискретный анализ". Вып. 7. Новосибирск, ИМ СО АН СССР. 1966. С. 3-11.

[2] Сенько О.В. Использование процедуры взвешенного голосования по системе базовых множеств в задачах прогнозирования// М. Наука, Ж. вычисл. матем. и матем. физ. 1995, т. 35, № 10, С. 1552-1563.

[3] V.V.Ryazanov, Recognition Algorithms Based on Local Optimality Criteria, Pattern Recognition and Image Analysis. 1994. Vol.4. no.2. pp. 98-109.

[4] Сенько О.В., Кузнецова А.В. Метод предварительной селекции признаков. // Докл. Всеросс. Конф. «Матем. методы распознавания образов - 11». М.: Регион-Холдинг, 2003, С. 171-173.

[5] Обухов А.С., Рязанов В.В. Применение релаксационных алгоритмов при оптимизации линейных решающих правил. Доклады 10-й Всероссийской конференции "Математические методы распознавания образов (ММРО-10)", Москва, 2001, 102-104.

[6] William H. Wolberg and O.L. Mangasarian: "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990, pp. 9193-9196.

[7] R.O. Duda and P.E. Hart. Pattern Recognition and Scene Analysis. A Wiley-Interscience Publication. John Wiley and Sons. New York, 1973.

[8] Ю.И.Журавлев, Об алгебраическом подходе к решению задач распознавания или классификации, Проблемы кибернетики. М.: Наука, 1978. Вып.33. С.5-68.

[9] Рязанов В.В., Челноков Ф.Б., О склеивании нейросетевых и комбинаторно-логических подходов в задачах распознавания, Доклады 10-й Всероссийской конференции "Математические методы распознавания образов (ММРО-10)", Москва, 2001,115-118.

[10] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition, Appeared in: Data Mining and Knowledge Discovery 2, 1998, 121-167.

[11] Yu.I. Zhuravlev, V.V. Ryazanov, O.V. Sen'ko, A.S. Biryukov, D.P. Vetrov, A.A. Dokukin, N.N. Katerinochkina, A.S. Obukhov, M.Yu. Romanov, I.V. Ryazanov, F.B. Chelnokov. The Program System for Data Analysis "Recognition" (LOREG). The 6th German-Russian Workshop "Pattern Recognition and Image Understanding". Proceedings. Novosibirsk, Russia, 2003, pp. 255-258.

[12] Ganster H., Gelautz M., Pinz A., Binder M., Pehamberger H., Bammer M., Krocza J. Initial Results of Automated Melanoma Recognition //Proceedings of the 9th Scandinavian Conference on Image Analysis, Uppsala, Sweden, June 1995, Vol.1, pp. 209-218.

[13] Dyukova Ye.V., Ryazanov V.V., The Solution of Applied Problems by Recognition Algorithms Based on the Voting Principle. VTs Akad. Nauk S.S.S.R., Moscow, 1986.

Author information

Alexander A. Dokukin – Dorodnicyn Computing Centre of the Russian Academy of Sciences, Vavilov st., 40, Moscow GSP-1, 119991, Russia; e-mail: dalex@ccas.ru

Oleg V. Senko - Dorodnicyn Computing Centre of the Russian Academy of Sciences, Vavilov st., 40, Moscow GSP-1, 119991, Russia; e-mail: senko@ccas.ru

TEACHING STRATEGIES AND ONTOLOGIES FOR E-LEARNING [13]

Tatiana Gavrilova, Michael Kurochkin, Victor Veremiev

Abstract. *The paper presents one approach aimed at developing teaching strategies based on the principles of ontological engineering. The research framework is targeted on development of methodology and technology that will scaffold the process of knowledge structuring for e-learning. The structuring procedure is the kernel of ontology development. Ontologies that describe the main concepts of the domains are used both for teaching and assessment techniques. Special stress is put on visual design as a powerful learning mindtool. The examples are taken from the courses on the foundations of artificial intelligence and intelligent systems development. These courses are delivered by the authors in St.Petersburg State Polytechnical University at School of Computer Science and in Poland in the First Independent University.*

Keywords*: E-learning, Ontologies, Visual Knowledge Engineering, Expert Systems Building Tools, Knowledge Acquisition, Knowledge Sharing and Reuse.*

1. Introduction

The drawback of e-learning is lack of feedback from the teacher or tutor. That is why the courseware should be more precisely structured that in face-to-face teaching.

The idea of using visual structuring of teaching information for better understanding is not new. Concept mapping [Sowa, 1994; Jonassen, 1998, Conlon, 1997] is scaffolding the process of teaching and learning for more than 20 years. Visual representation of the general domain concepts is facilitative and helps both learning and teaching. A teacher now has to work as knowledge analyst or knowledge engineer making the skeleton of the studied discipline visible and showing the domain's conceptual structure. This structure is now called "ontology". However, ontology-based approach is rather young. It was born in knowledge engineering [Boose, 1990; Wielinga, Schreiber, Breuker, 1992], then it was transferred to knowledge management [Fensel, 2001].

The short prehistory of knowledge engineering (KE) techniques and tools (including knowledge acquisition, conceptual structuring and representation models), the overall overview of which is presented in [Adeli, 1994; Scott, Clayton, Gibson, 1994], is an ascending way to the development of the methodology that can bridge a gap between the remarkable capacity of human brain as a knowledge store and the efforts of knowledge engineers to materialise this compiled experience of specialists in their domain of skill.

Beginning from the first steps to nowadays knowledge analysts have been slightly guided by cognitive science. So major part of KE methodology suffer of fragmentation, incoherence and shallowness.

The last years the main interest of the researchers in this field is concerned with the special tools that help knowledge capture and structuring. This generation of tools is concerned with visual knowledge mapping to facilitate knowledge sharing and reuse [Eisenstadt, Domingue, Rajan, Motta, 1990; Tu, Eriksson,. Gennari, Shahar,. Musen, 1995; Johnassen, 1998]. The problem has been partially solved by developing of knowledge repositories called ontology servers where reusable static domain knowledge is stored (e.g. projects as Ontolingva, Ontobroker, KA2, etc.)

In tutoring systems teachers are supposed to reuse the domain ontologies in order to support the description of the discipline they taught and the problem-solving methods of their domain. The idea is to allow teachers to model both domain and problem-solving knowledge using the same visual language. Ontology design also may be used as an assessment procedure. Students show their knowledge and understanding while creating ontologies.

Knowledge entities that represent static knowledge of the domain are stored in the hierarchical order in the knowledge repository and can be reused by other teachers. At the same time those knowledge entities can be also reused in description of the properties or arguments of methods of another knowledge entity. Concept maps modelling language that is designed in the framework of the described project is based on a class-based object-

[13] The work is partially suported by Russian Foundation for Basic Studies (grant 02-01-00466)

oriented language which is aimed to support typing and parameterisation of knowledge entities. Due to the class subsumption and polymorphism of classes the reasoning process becomes extremely flexible. This non-formal system allows to reason on a large set of knowledge and to apply problem-solving rules described for the higher level knowledge entities to the lower level knowledge entities based on the class inheritance. In contradistinction to ontology server approach where static knowledge described is very specific to the domain, the approach which is taken in the paper simplifies reusability of the dynamic knowledge and as a consequence building of large-scale knowledge bases with a flexible reasoning capability.

The proposed ideas and methods may be applied to those tutoring systems where general understanding is more important that factual details. We used such approach in teaching Artificial Intelligence, Neuroscience and Computer Graphics.

2. Ontological Engineering

An ontology is a set of distinctions we make in understanding and viewing the world. There are a lot of definitions of this milestone term [Neches et al,1991; Gruber, 1993; Guarino et al, 1995; Gomez-Peres, 1999]:

1. Ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary.
2. Ontology is an explicit specification of a conceptualization.
3. Ontology as a specification of a conceptualization.
4. Ontology as an informal conceptual system.
5. Ontology as a formal semantic account.
6. Ontology as the structured vocabulary.
7. Ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.

All these definitions together clarify the ontological approach to knowledge structuring on one hand, on the other hand give enough freedom to the open thinking. So ontological engineering gives the intuitively clear representation of company structure, staff, products and relationship among them.

Mane researchers and practitioners argue about distinctions between ontology and user's conceptual model. We supposed that ontology corresponds to the analyst's view of the conceptual model, but is not the model itself.

Ontologies are useful structuring tools, in that they provide an organising axis along which every student can mentally mark his vision in the information hyper-space of domain knowledge. Rather often we can't express all the information in one ontology, so subject knowledge storage includes a set of ontologies. Some problem may occur when jumping from one ontological space to another. But constructing of meta-ontologies may help.

Ontology development also faces the knowledge acquisition bottleneck problem. The ontology developer comes up against the additional problem of not having any sufficiently tested and generalised methodologies recommending what activities to perform and at what stage of the ontology development process these activities should be performed. That is, each development team usually follows their own set of principles, design criteria and steps in the ontology development process. The absence of structured guidelines and methods hinders the development of shared and consensual ontologies within and between teams, the extension of a given ontology by others and its reuse in other ontologies and final applications [Guarino, Giaretta, 1998].

Till now, few domain-independent methodological approaches have been reported for building ontologies [Swartout, Patil, Knight, Russ, 1997; Fensel, 2000]. These methodologies have in common that they start from the identification of the purpose of the ontology and the need for domain knowledge acquisition. However, having acquired a significant amount of knowledge, major researchers propose a formal language expressing the idea as a set of intermediate representations and then generating the ontology using translators. These representations bridge the gap between how people see a domain and the languages in which ontologies are formalised. The conceptual models are implicit in the implementation codes. A reengineering process is usually required to make the conceptual models explicit. Ontological commitments and design criteria are implicit in the ontology code.

This paper proposes the most explicit way for ontology design - to use the visual representation in a form of a tree or set of trees.

Figures 1 and 2 illustrate the idea how ontology bridge the gap between chaos of unstructured data (names of different models and techniques for knowledge representation) and clear knowledge of modern classification. Our approach shows that ontology development process needs some creative efforts of meta-concepts definition that helps to name the groups and stucture the chaos.

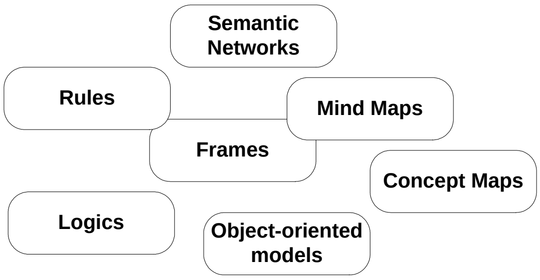


Figure 1. Unstructured set of Knowledge Representation Models

Ontology developers (who are unfamiliar with or simply inexperienced in the languages in which ontologies are coded, e.g. DAML, OIL, RDF) may find it difficult to understand implemented ontologies or even to build a new ontology.

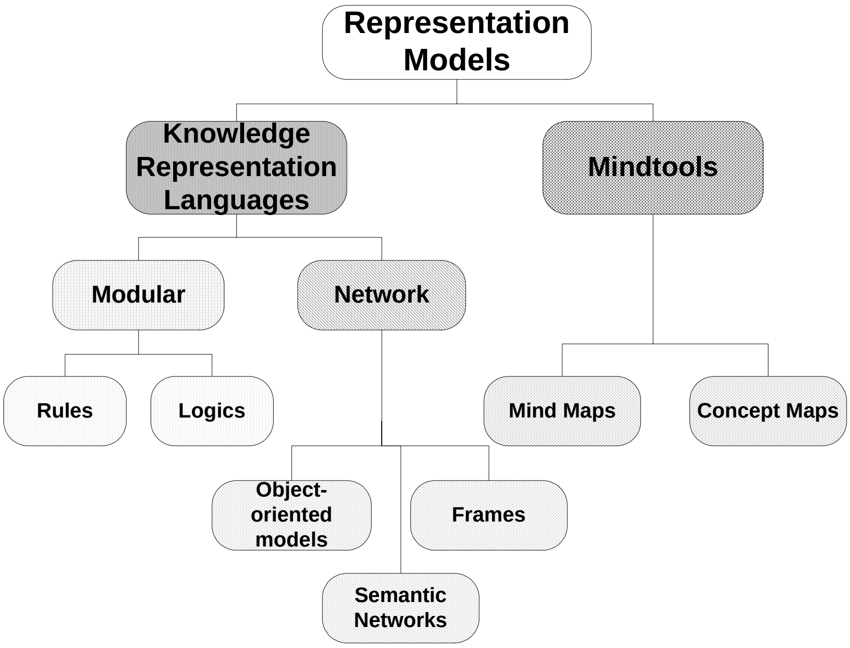


Figure 2. Ontology “Knowledge Representation Models”

It is easier for any educationalist simply to draw the ontology using well-known to everybody “pen and pencil” technique then to study these languages.

3. Object-structured approach

Although there are some methods that are rather powerful and versatile [Kremer, 1998], the teacher as knowledge analyst is still weakly supported while working for a set of ontologies to describe the main subject knowledge. This process is the most important and critical stage in the courseware preparation life cycle - transition from elicitation to conceptualisation by understanding and realisation of the subject information structure and main reasoning way. The teacher may do this sophisticated g procedure alone or to ask for help of professional analysts.

In this way, a special methodology Object-Structered Analysis (OSA) has been developed [Gavrilova and Voinov, 1992-2000], which is intended to help knowledge analyst to perform the most informal step of knowledge acquisition, concluding in prior conceptual structuring of the subject domain. The approach presents the enhancement of classical structured analysis methodology [Sowa, 1994; Yourdon, 1990] to knowledge engineering.

OSA is based on decomposition of subject domain into several (3-8) strata (Tab.1). The number of strata is considered by the analyst. This multi-step and time-consuming procedure is methodological base for effective constructing of subject ontologies.

s1	***WHAT FOR*** Knowledge	*Strategic* Analysis: Targets, Aims, Requirements, Constraints.
s2	***WHO*** Knowledge	*Organisational or Historical* Analysis: Main Researchers, Human Resources, Actors.
s3	***WHAT*** Knowledg**e**	*Conceptual* Analysis: Main Concepts, Processes, Entities and Relationships between them.
s4	**HOW TO** Knowledge	*Functional Analysis*: Main Algorithms, Decision Procedures, Business Processes Modelling, Decision Making Models.
s5	***WHERE*** Knowledge	*Spatial* Analysis: Geography, Environment, Communications, etc.
s6	***WHEN*** Knowledge	*Temporal* Analysis: Historical Dates, Schedules, Time Constraints, etc.
s7	***WHY*** Knowledge	*Causal* Analysis: Explanations to Decision Making Models.
s8	***HOW MUCH*** Knowledge	*Economical* Analysis: Resources, Losses, Incomes, Revenues, SWAT, etc.

Table 1. Matrix for OSA

Level → **Stratum↓**	**Domain Level in general (u_1)**	**Problem Level (u_2)**	**Sub-Problem Level (u_3)**	**……**	**(u_n)**
Strategic Analysis s_1	E_{11}	E_{21}	E_{31}	E_{i1}	E_{n1}
Organisational Analysis s_2	E_{21}				
Conceptual Analysis s_3	E_{31}				
Functional Analysis s_4	E_{41}				
Spatial Analysis s_5	E_{51}				
Temporal Analysis s_6	E_{61}				
Causal Analysis s_7	E_{71}				
Economical Analysis s_8	E_{81}				
…..				E_{ij}	
s_m	E_{m1}				E_{mn}

Filling that matrix is performed into two steps:

Step 1. Global (vertical) analysis, i.e. decomposition of the heterogeneous domain information into the groups related to mentioned above methodological strata.

Step 2. Local analysis of each individual stratum (horizontal), concluding in maintenance of gradually detailed structures. The number of levels depends on peculiarities of the subject domain and could vary dramatically for different strata. From the point of view of methodology the number of levels n<3 indicates ill-structured domain knowledge.

The first level (or column 2 in the table) corresponds to the discipline information as a whole. The second one corresponds to the problem that is studied now. The others may correspond to particular sub-problems, depending on the required reasonable deepness of detailing. The procedure of the described analysis may be performed both in top-down and bottom-up strategies, including their possible mixture.

The formation of strata with more of less definite meaning as described in Tab.1 allows to avoid many traditional didactic mistakes in teaching and learning. The minimal obligatory set of strata for the course structuring development is:

s3: Conceptual Structure or subject ontology.

s4: Functional Structure or main problem solving procedures.

Other strata are designed and developed if needed by subject peculiarities, e.g. spatial and temporal analysis strata (s5 and s6) may be formed in those disciplines which study construction or management where the issues of scheduling, real-time operations, real object manipulation are substantial.

Step 1 algorithms may be sketched in such form:

1.1: Gather all the data and knowledge of discipline identification

1.2: Select a set of N strata to be formed ($N \geq 3$).

1.3: For each i-th stratum select a subset of all available information, relevant to that stratum and represent it in way appropriate to that stratum (see below).

1.4: If there remains unused bulk of information, increase number of strata and repeat step 1.3. Otherwise, begin the horizontal analysis of each declared stratum.

Step 2 is horizontal analysis of strata that depends on the number of columns in OSA matrix and may be performed in two ways: deductive (top-down) and/or inductive(bottom-up). As the most essential stratum is s3 (WHAT-analysis), the horizontal analysis for it is concluded by resulting conceptual structure or a set of the domain ontologies.

Analogous algorithms were developed and practically tested and evaluated by the authors during developing of distance learning courses for different branches of computational science and for artificial intelligence (AI).

4. Teaching Ontologies in Artificial Intelligence

We have developed more than 20 teaching ontologies (What-knowledge conceptual structures s3) helping to understand and to remember main concepts of AI. Fig.3 shows one of them (it includes a part of Fig.2)..

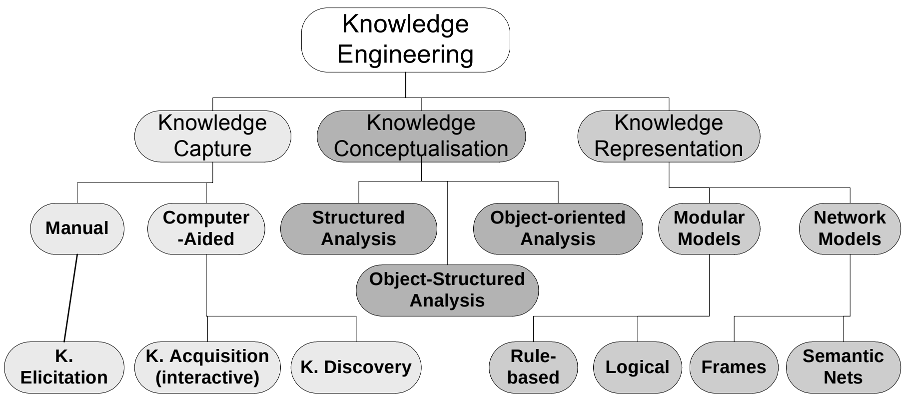


Figure 3. Ontology "Knowledge Engineering"

We worked out several tips to add expressiveness to the ontology on the design stage.

1. Use different font sizes for different strata
2. Use different colours to distinct the subset or branch
3. Use vertical layout of the tree
4. If needed use different shapes of nodes

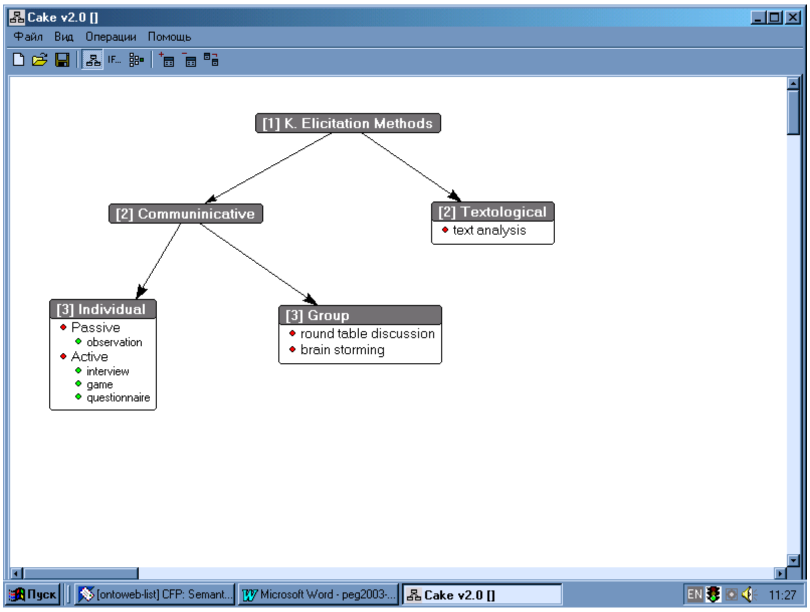


Figure 4. Screenshot with a part of ontology "Knowledge elicitation techniques".

It is possible to use any of graphical editors to design ontology, e.g. PaintBrush, Visio, Inspiration. But computer program, which could be really useful for a knowledge engineer on the described stages of structuring of the subject domain, should necessarily follow the phenomenological nature of the knowledge elicitation and described above algorithms. This program must not frustrate the knowledge engineer with any "game rules" which were not evident for him/her. Ideally, it should adjust itself for particular cognitive features of the knowledge engineer. Moreover, each of the stages of analysis described above may be represented visually in its proper terms, as is already approached in some commercial expert system shells.

A special visual tool was developed and named CAKE-2 (Computer Aided Knowledge Engineering by leading programmer Tim Geleverya, previous release by Alex Voinov). CAKE illustrates the idea of knowledge mappability, that find another application in the data mining and structuring for heterogeneous data base design. Its first prototype is described briefly in [Gavrilova, Voinov, 1996]. CAKE-2 proposes a kind of a visual knowledge representation language, which analogues may be found in a wide range of visual software construction tools – from large CASE's to Visual Basic. In particular, it supports the principle of a bi-directional mutually unambiguous correspondence between the two-dimensional visual object description syntax with the traditional one-dimensional one.

CAKE-2 is based on classical structured analysis methodology [Yourdon, 1989] enriched by new results that gives a teacher the opportunity to use special graphical interface to create ontology, to save it and to compile into the knowledge base (if needed).

Fig.4 presents CAKE's screenshot with fragment of the ontology of knowledge elicitation methods.

5. Discussion

Our approach puts stress on the knowledge structuring for better understanding of main course ideas in e-learning. The use of visual paradigm to represent and support the teaching process not only helps a professional tutor to concentrate on the problem rather than on details, but also enables pupils and students to process and understand great volume of information.

A better apprehension of teaching information might be achieved by imposing a knowledge structure on it. This may improve later usage of this information, comparing, generalisation, and so on. Therefore, a visual knowledge structure editor plays here a role of a two-dimensional, pictorial conspectus of the regarded piece of information.

Acknowledgements

The authors are very grateful to Tim Geleverya who assisted in the described research.

References

Adeli H. (1994) *Knowledge Engineering.* McGraw-Hill. New-York.

Boose, J.H.(1990) Knowledge Acquisition Tools, Methods and Mediating Representations. In: Motoda, H., Mizogochi, R., Boose, J., Gaines, B. (Eds.): *Knowledge Acquisition for Knowledge-Based Systems,* IOS Press, Ohinsha Ltd., Tokyo,123-168.

Conlon T. (1997) Towards Diversity: Advancing Knowledge-based Modelling with Knowledge Acquisition. In *Proceedings of 8th International PEG Conference.* Sozopol, Bulgaria.,379-386.

Eisenstadt M., Domingue J., Rajan T., Motta E. (1990) Visual Knowledge Engineering. *In IEEE Transactions on Software Engineering*, Vol.16, No.10, 1164-1177.

Fensel D. (2001) *Ontologies: A Silver Bullet foe Knowledge Management and Electronic Commerce*. Springer.

Gavrilova T.A., Voinov A. (1996) Visualized Conceptual Structuring for Heterogeneous Knowledge Acquisition // Proceedings of *Int. Conf. on Educational Multimedia and Hypermedia EDMEDIA'96, MIT, Boston, USA*, 258-264.

Gomez-Perez A. Tutorial on Ontological Engineering. *Int. Conference on artificial Intelligence IJCAI'99*.

Gruber, T. (1993) A translation Approach to portable ontology specifications. *Knowledge Acquisition*. Vol. 5, 199- 220.

Guarino, N., Giaretta, P. (1998) Ontologies and Knowledge Bases: Towards a Terminological Clarification. In *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*. IOS Press, 25- 32.

Jonassen, D.H. (1998). Designing constructivist learning environments. In C.M. Reigeluth (Ed.), *Instructional design models and strategies*, 2nd ed. Mahwah, NJ: Lawrence Eribaum.

Kremer R. (1998) Visual Languages for Knowledge Representation. In Proceedings of KAW'98 *Eleventh Workshop on Knowledge Acquisition, Modeling and Management*, Banff, Alberta, Canada, 124-129.

Neches et al, (1991) Enabling Technology for Knowledge Sharing. *AI Magazin,*. Winter, 36- 56.

Sowa, J. F. (1984) *Conceptual Structures: Information Processing in Mind and Machine*. Reading, Massachusetts, Addison-Wesley.

Scott A., Clayton J.E., Gibson E.L. (1994) *A practical Guide to Knowledge Acquisition*, Addison-Wesley Publ.

Swartout B., Patil R., Knight K., Russ T. (1997) Toward Distributed Use of Large- Scale Ontologies Ontological Engineering. *AAAI- 97 Spring Symposium Series,* 138- 148.

Tu S., Eriksson H., Gennari J., Shahar Y., Musen M. (1995). Ontology-Based Configuration of Problem-Solving Methods and Generation of Knowledge-Acquisition Tools: Application of PROTEGE-II to Protocol-Based Decision Support. *Artificial Intelligence in Medicine,* N7, 257-289.

Wielinga B., Schreiber G., Breuker J. (1992) A Modelling Approach to Knowledge Engineering. *Knowledge Acquisition*, 4 (1). Special Issue, 23-39.

Yourdon, E. (1989) Modern Structured Analysis, Prentice-Hall Bristol, Computer Science Department.

Authors' Information

Tatiana Gavrilova, Michael Kurochkin, Victor Veremiev:

Saint-Petersburg State Polytechnical University; Intelligent Computer Technologies Dpt.

Politechnicheskaya 29/9, St. Petersburg - 195251, Russia, Tel: + 07 812 371 82 72, Fax: + 07 812 373 06 89; gavr@limtu.spb.su

QUALITY AND EFFICIENCY IMPROVEMENT OF THE TEACHING PROCESS OF MATHEMATICS BY APPLICATION OF SPECIALIZED SOFTWARE

Tsvetanka Kovacheva

Abstract. *The paper presents application of educational technologies in teaching process and their contribution to the improvement of the quality and efficiency of students training. The need for training of highly qualified specialists, who can solve application problems creatively and striving for self-perfection, continuously increases the requirements to the knowledge and skills of the students of the Technical Universities (TUs). They should be acquainted also with the new technologies for computer mathematical modeling. All these change the requirements to the quality of the mathematical education in TUs and require application of specialized software in the training process.*

1. Introduction

The informational blow-up in almost all spheres of the human activities caused a sharp increase of the educational level of the students and reformation of the educational system as a whole. One of these new directions is the application of new Educational Technologies (ET). They have two basic *components*: *technologies within the teaching (media),* which describe the various resources in the education and *technologies for education*, which include planning, organization, conduction and evaluation of the whole educational process. Improvement of the *education quality* is connected with use and perfection of its components, namely: quality of curricula and syllabus, books of mathematical problems, textbooks, quality of educative work, application of specialized software, level of research with inclusion of students, status of the material-technical base of the university, infrastructure development, discipline, etc.

Choice of appropriate media is necessary in order to reach the objectives of the education [5,6]. The main factors for selection include immediate objectives and time, media access, human resources, costs and expenses. The computer-based resources prove to be feasible for teaching in mathematics of non-mathematical specialities, due to the opportunities they offer. They allow storage of huge quantities of information, provide high speed and flexible operation with the information. They allow interactive training, provide larger quantity and great diversity of examples, better opportunities for analysis, visualization, combination of various strategies of teaching, depending on the educational objectives. The use of computers and Application Software Packages (ASP) for automation of mathematical calculations in the teaching process, definitely contribute to a better and faster mastering of the presented material. The computational and graphical features of ASP render a significant support to improve the quality of teaching mathematics. They promote organizational betterment and efficiency of the teaching process and support the students to extend their knowledge in many tasks and problems of various chapters of the higher mathematics.

The efficiency of the teaching process is a function of the human capacity and ET. For this reason, each teacher strives to be more helpful to his students, encouraging them to efficient training, supports them to acquire habits and skills to learn continuously and directs them to ratiocination, research and uncovering of the scientific truth.

2. Application of Specialized Software in the Teaching Process of Mathematics

The improvement of the material-technical base of TU – Varna allowed establishment of laboratories at respective departments, which are equipped with computers and other technical facilities. Opportunities were created for application of new ET for quality and efficiency improvement of the learning process of mathematics in the University. This allowed conduction of laboratory exercises (LE) on some chapters of mathematics – Mathematical Analysis – Part 2 (MA2), Selected Chapters of Mathematics (SCM), Probability Theory and Statistics (PTS), Numerical Methods (NM) with use of available ASP. The exercises are envisaged for first and second grade students of almost all specialties of the University.

There are many mathematical ASP for symbolic (analytical) mathematical transformations and numerical methods, such as: Mathematika, Maple, Matlab, MathCad, Excel, etc. The following ASP are used for LE on mathematics:

- ✓ Maple – for LE on MA2, SCM and NM for engineering specialties;
- ✓ Matlab – LE on Mathematics for Electronics specialty (for the new Academic year)
- ✓ Excel – LE on PTS for specialties: Public Administration, Economic Management, Ecology

Maple is frequently used in research [1,2,4]. It allows conversion of documents into LaTeX format. Maple possesses a well-developed programming language. This allows creation of instructions (commands) for solution of specific tasks. The package is realized for all environments - Windows, Macintosh, Linux, UNIX. Provides determination of limits, derivatives, integrals, finite and infinite sums, products. Algebraic systems of equations and inequalities can receive symbolic and numerical solution. It is possible to obtain roots of equations, to solve analytically and numerically systems of ordinary differential equations and some classes of partial differential equations. The good text editor and the excellent graphic capabilities allow professional presentation of projects. The interface supports the worksheet concept, which integrate input instructions, output and graphics in one and the same document. It allows simultaneous operation with several worksheets, establishing dynamic links among them and the calculations can be transferred from one worksheet to another. It is possible to launch several programs simultaneously which provides possibility for comparison of the calculations at various initial values of the variables.

Matlab is a peculiar high level programming language, oriented for engineering and scientific tasks [1,3]. It allows storage of the documents in C language format. Matlab supports mathematical calculations, scientific graphic visualization and programming. The Matlab system is most frequently used for calculations, development of algorithms, numerical experiments, simulation modeling, breadboarding and prototyping, data analysis, study and visualization of results, scientific and engineering graphics. The tasks for extension of the system are solved in Matlab by specialized packages – Toolbox. One of them – Simulink, has a special position – it is package for simulation and analyses of dynamic systems and units. It allows study of various systems for technical, physical and other applications. It is used in all stages of simulation and significantly facilitates the work. Matlab includes a large library of elements (blocks) for analysis of linear and nonlinear systems, discrete, continuous and hybrid processes. There are several integrators for solution of ordinary differential equations, a special accelerator for calculations and it is possible to create and compile modules written in C.

The spreadsheets *Excel* are one of the most popular office packages for work with tabulated data [7]. They provide not only simple arithmetic calculations with groups of cells from the spreadsheets, but also allow to carry out sufficiently sophisticated calculations and data analysis with the help of the so-called “add-ins”. Excel solves problems of analytical geometry, linear algebra, optimization analysis, PTS.

A trend of rapprochement and integration of various packages is observed nowadays. The last versions of Matlab include library for analytical transformations Maple, while Maple allows calls to some functions of Matlab and Excel.

The packages are convenient for training students, as they do not require preliminary computer experience. They possess user friendly interface and are easy for mastering. The interaction with the user is fast and it is possible to cover a large part of the mathematics syllabus. This is the main reason why these packages are selected in TU for teaching students in mathematics.

3. Main Tasks of LE on mathematics

The objective of the laboratory exercises is to acquaint students with the capacity of the used packages for solution of basic problems of the studied chapter of mathematics, while extending and consolidating the knowledge accumulated during lectures and seminar exercises [4]. Seven exercises are conducted for each subject. A correct selection of the tuition materials and optimization of the LE are necessary.

During LE on MA2, students learn to determine partial derivatives and extremes of functions of two and more variables, to expand in series these functions, to solve analytically and numerically ordinary differential equations and to compare the solutions, to calculate double and triple integrals. They are also acquainted with the graphic capacities of the packages. The geometrical interpretation of the introduced mathematical notions – extremum of function, directional field and integral curves of differential equation, etc., as well as graphs of functions under consideration, double and triple integrals domains, etc., allow students to figure out visually and to make sense of the teaching material and to get incentives for more thorough learning.

The main tasks of students during LE on SCM are connected with operations with complex numbers and elementary functions of a complex variable, expansion of functions in Fourier series, analytical and numerical solution of partial differential equations, as well as PTS problems.

Particularly useful for students are the exercises for familiarization with and application of various numerical methods for solution of specific problems. LE, introduced for some specialties of TU for NM, consider the solution of systems of equations (simple iteration method, method of successive approximations and Seidel method) for nonlinear equations and systems of nonlinear equations (iteration method and Newton method), numerical differentiation and integration, function fitting and approximation (interpolation polynomial of Lagrange and Newton).

During LE on PTS, students are acquainted with some elements of the combinatorial calculus and their application to solve problems of probability theory, finding distribution laws and numerical characteristics of discrete and continuous variables as well as processing of small and large samples. Students gain skills to apply statistical methods, to calculate statistical characteristics of data arrays, to approximate data by given dependencies, to generate random variables, to evaluate numerically statistical functions, to introduce various data transformations and to make graphic presentations.

4. Conclusions

The use of computers and ASP is of significant importance to improve the qualification of the lecturers and their self-education. This supports them to prepare and conduct efficiently the scheduled LE. By creative use of the capacities of the packages, they could make appropriate specifications to lead exercises aiming at more thorough presentation of the considered topic with active participation of students.

In order to establish habits for use of the mathematical simulation methods, some examples of mathematical models for specific problems can be presented, depending on the student's specialty. For instance – use of ordinary differential equations to solve problems in electrical engineering, economics, ecology and partial differential equations for problems of mathematical physics.

To increase the interests of students to LE and to enrich their information culture it is good idea to grant them the used product, various training and testing programs, useful and interesting Internet links for electronic textbooks, practical examples and other relevant aids.

References

1. Govoruhin V., Tsibulin V "Computer in Mathematical Studies: Maple, MATLAB, LaTeX. Training Course", Peter, 2001. (in Russian)
2. "Maple Application Center", http://www.mapleapps.com/
3. "Consultancy Center Matlab of Softline Co." - http://www.matlab.ru
4. Ts. Kovacheva "Application of Maple for Mathematics Teaching in Universities" Workshop: "Teaching in Mathematics – Traditions and Prospects", Varna, 2003.
5. Information Technologies in Science and Education: - http://osu.tpu.ru/edu/tech/
6. Technology of Educational Resources: http://www.webuniversity.ru/
7. Gelman V. "Solution of mathematical problems with means of Excel", Moscow, "Piter", 2003.

Author Information

Tsvetanka Kovacheva - Department Mathematics of Technical University, 9010 Varna, 1 Studentska str., e-mail: ts_kovacheva@hotmail.com

MUSIC AS THE SOURCE OF INFORMATION INFLUENCE AND SOUL EDUCATION

L. Kuzemina

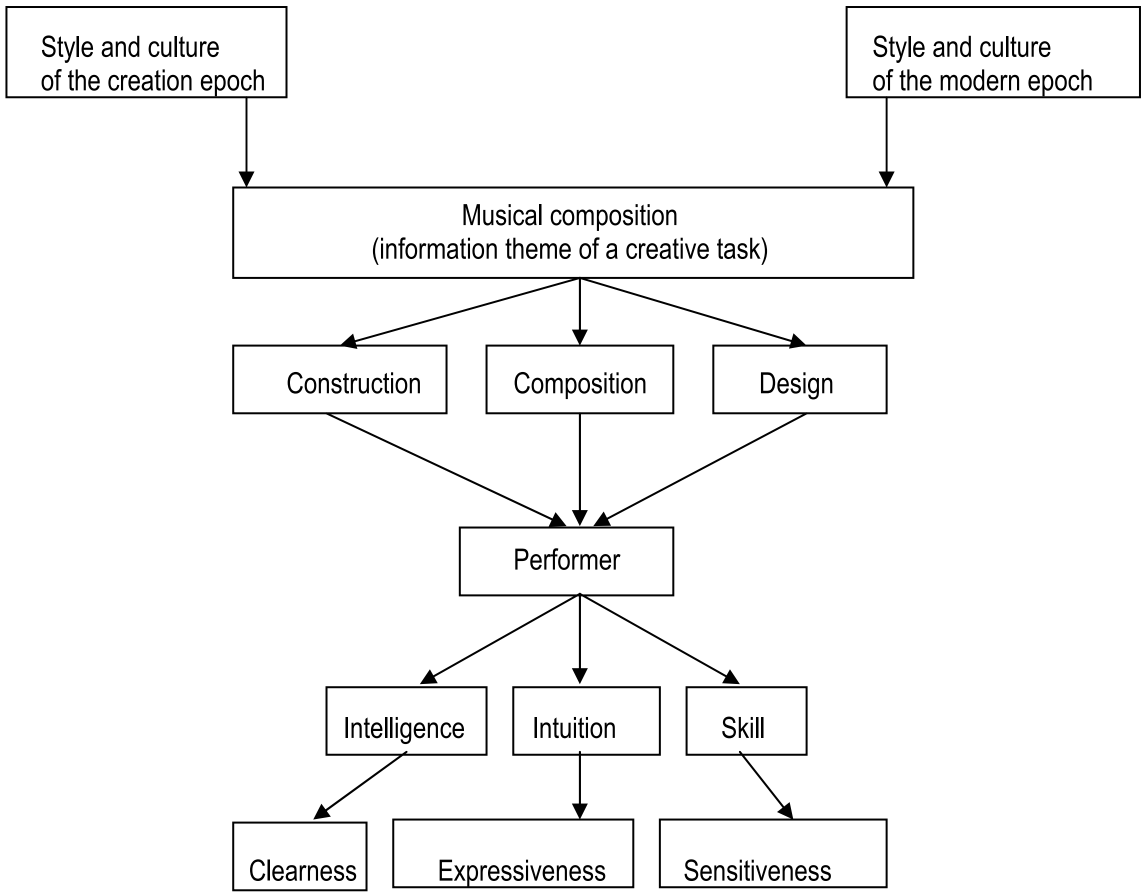
Extended Abstract

Unlike other works of art (painting, sculpture, etc.) a musical composition should be performed, it should sound to become accessible. Therefore, the role of the musical masterly performance is extremely important. But presently it has increased in importance when music through mass communication media i.e. radio, television, sound recording becomes in the full sense of the word the property of millions.

The educative importance of music as a means of forming the spiritual culture of a person grows as well. Clearly, a great deal depends on performers. They can breath new life into the musical composition, but they can necrotize it. They can educate with music, they can awaken love in it, and, vice versa, discourage a person from interest, love for music etc.

The system analysis applied to studying music ensures understanding of a musical composition as an information source reflecting the historical heritage of mankind. Informational language used by an author for ciphering human feelings, people relations, events etc., can be presented by a performer at another time of the society existence taking into account its culture and technique of performance for the audience of another epoch.

Systematically the process of revealing the musical composition “biography” can be presented in the following way:



This scheme clearly shows the way from an author to performer and then to a hearer. A skill to unite an accumulated information knowledge (sociological, historical, musical theory and contiguity arts) in the work on the informative presentation of a musical composition must be “directed” to the understanding of a musical composition from the smallest units to the whole by a performer, reveal the role of each episode in the events development, in comparison of the image-bearing spheres (composition) and find such methods and information means of performer’s expression which corresponded most of all to the pithy side of music (design).

When solving the set problems of the musical composition system analysis the so called linguistic uncertainty emerges.

Linguistic uncertainty emerges due to the fact that a person is capable to “simplify” descriptions and requirements, this makes him possible to realize the efficient presentation of a musical composition which is connected with the use of such indefinite and fuzzy concepts as: “much”, “little”, “good”, “bad” etc. The use of such concepts when building models of a musical composition became possible due to introduction of the concepts: ”fuzzy set”, “fuzzy variable” and “linguistic variable” into consideration.

There are many approaches to the solution of the decision-making problems, which are acceptable in the given case. Formally, the problem can be formulated in the following way:

1. There is a set of possible presentations of a musical composition, which can be chosen by a performer.
2. There is a lot of possible alternatives.
3. It is required, having studied the performer’s preferences system, to build a model of the alternatives’ choice being the best to some extent.

The basis for the set problem solution can be the results of investigations performed by Yavorsky.. Theoretical concepts of Yavorsky are distinguished not only for a deep scientific substantiation, but for interconnection. He is “the first Russian scientist who created the generalized system embracing all musical speech as a whole”.

The natural extension of the theoretical searches was the work of Yavorsky in the field of the musical education. How one can educate a creative person in art and not a craftsman in the field of the musical art? How can one train a widely educated, creatively thinking, socially active expert? Yavorsky made a considerable contribution to these problems solution. His theory was approved in all the links of the three-level system of the musical education, the idea of its creation belonged just to him. Even the first experiments in its application gave positive results: the application of theoretical concepts to the special classes in piano stimulated development of thinking, creative inclinations, essential qualities for music, trained to a conscious approach to music, to ability to join musical composition elements into a single whole. Synthesis of understanding, acoustic and motor bases made it necessary further promotion of this theory in the musical pedagogy. Education in the First Moscow State Musical College was based on Yavorsky’s theoretical system. The pedagogical principles of the scientist were grounded on it. It is well known that it was exactly Yavorsky who revised the educational programs and created new educational plans including a wide complex of subjects. They encompassed the study of philosophy, history, sociology, psychology, logic, pedagogy, physiology, information on all trends in art (poetry, literature, fine arts, theatre, plastic art etc.). Yavorsky’s method as an integral system of training musicians didn’t recognise absolutely the subjects’ isolation, their alienation. The knowledge complex interconnection, interaction widened horizons in the field of art itself for conscious activity in the field of a wide social life. An element of creative work should be included in every subject, in the programs of all courses, as Yavorsky stressed “…everywhere schools should tech not only to read the written, but to pronounce own words”.

Authors' Information

Kuzemina L.A.: Prof. of Kharkiv State Institute of Arts, kuzy@kture.kharkov.ua

FORMAL DEFINITION OF THE CONCEPT "INFOS"

K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova

Abstract: *The concept INFOS is very important for understanding the information phenomena. Because of this, it is basic for the General Information Theory. The more precise formal definition of this concept is given in the paper.*

Introduction

The genesis of the concept of **Infos** started from the understanding that the concept **"Information Subject"** is perceived as human characteristic. It is clear that in the nature there exist many creatures which may be classified to this category. To exclude the misunderstandings we decide to introduce new word to denote all possessors of the characteristics of the Information Subject. This word is **"*INFOS*"**.

The concept **"Infos"** is basic for the General Information Theory [Markov et al. 2003]. Its definition is only the starting point for further investigations and building the ***Infos Theory***.

The variety of types of Infoses in the real world needs to be investigated and classified in the future research. At the first step, we may propose that may be at least two main types of Infoses exist:

- ***infogens*** - the natural creatures;
- ***infotrons*** - the artificial creatures.

Also, the Infos Theory needs to give answers to many other very important questions, such as:

- What is the nature of the activity of the Infos?
- What is the difference between the living level of the Infos and the non-living one?
- Is it true that the boundary between non-living and living entities is self-reflection and internal activity for satisfying the secondary (information) possibilities for internal or external contact?

Etc.

It is impossible to answer to all questions in a single article. We may make only the next little step. This is the aim of the present paper.

The formal definition of the concept of Infos needs theoretical basis which may support further mathematical investigations in this area. The first attempt was made in [Markov et al., 2003]. This brochure is the first attempt to compound all definitions of the main concepts of the General Information Theory, arisen during the years.

In [Markov et al., 2003a] we have continued our work with some new results, which make the previous formal definitions more precise. The main concepts, defined in this article are: "Entity", "Impact", "Interaction", "Reflection", and "Information".

The present work extends the basic definitions of [Markov et al., 2003a] with formal definitions, which concern the concepts "Information Witness", "Activity", and "Infos".

The Information Witness

Let's remember the *definition* of "Entity" [Markov et al., 2003a].

The **entity** A is the couple $A=(E_A,R_A)$ where:

E_A is a collection of sub-sets of a set M_A;

$R_A=\{r_i|i\in I, \text{ I is a set}\}$ is a nonempty set of relations in E_A, i.e.

$r_i\subset E_A\times E_A=\{(X,Y)|X,Y\in E_A\}$ is a relation and $\check{r}_i=r_i\cup\{(X,Y)|(Y,X)\in r_i)\}$, $\forall i\in I$; and:

1. $\varnothing\in E_A$;
2. $M_A=\cup X,\ X\in E_A$;
3. $\forall r\in R_A$ and $\forall X,Y\in E_A \Rightarrow ((\exists(X,Y)\in\check{r})$ or

$$(\exists\ Z_1,..,Z_p\in E_A,\ Z_k\neq\varnothing,k=1,..,p : (X,Z_1)\in\check{r},(Z_1,Z_2)\in\check{r},\ldots,(Z_p,Y)\in\check{r})\ \square$$

Definition 1. Every relation $r_i \in R_A$ from the set of relations R_A is said to be a **state** of $A=(E_A,R_A)$.■

Let's mention that often the set M_A and collection E_A are compounded by entities.

In the entities with very large and complicated structure which are built by another entities from lower levels, the impact may be considered as a basis for relations (X,Y) which may be considered as a part of any state of the whole entity (relation $r_i \in R_A$).

Building the relationship between the entities is a result of the **contact** among them. During the contact, one entity **impacts** on the other entity and vice versa. In some cases the opposite impact may not exist, but, in general, the contact may be considered as two mutually opposite impacts which occur in the same time.

In [Markov et al., 2003a] we have formally defined the concepts "impact", "reflection", and "self-reflection".

Definition 2. Let $B=(E_B,R_B)$ is an entity. We say that B has **possibility for reflection**, if there exists any entity $A=(E_A,R_A)$ and $\psi_f=(A \rightarrow B)_\psi$ is an direct impact of A on B, and state $g \in R_B$ of B, which contain the reflection F_ψ.■

Definition 3. Let $A=(E_A,R_A)$ is an entity. We say that A has **possibility for self-reflection**, if there exists any entity $B=(E_B,R_B)$ and transitive self-impact $\xi(A \rightarrow B \rightarrow A)$ and state $h \in R_A$ of A which contains the reflection F_ξ.■

Let's remember the *definition* of Information [Markov et al., 2003a].

Let:

- $A=(E_A,R_A)$ and $B=(E_B,R_B)$;
- τ is an impact of A on B, i.e. $\tau=(A \rightarrow B)_\tau$, $\tau \in \Omega_{AB}$;
- $\exists$ entity $C=(E_C,R_C)$: $C \neq A$, $C \neq B$;
- $\exists$ $\psi=(B \rightarrow C)_\psi$ which can be composed with $\tau=(A \rightarrow B)_\tau$;
- $\exists$ transitive impact $\xi=\{\tau,\psi\}=(A \rightarrow B \rightarrow C)_\xi$;
- $\exists$ impact $\varphi=(A \rightarrow C)_\varphi$, $\varphi \in \Omega_{AC}$ and $\varphi \neq \xi$;
- F_φ is a reflection of the impact φ and F_ξ is a reflection of the impact ξ.

F_τ is **information** for A in B if $\exists r \in R_C:(F_\varphi \rightarrow F_\xi)_r$. □

The entity A is called **source**, the entity B is called **recipient**. The relation $r \in R_C$ for which $(F_\varphi \rightarrow F_\xi)_r$ is called **reflection evidence** and the entity C is called **information evidence**.

Let denote the information F_τ for A in B with information evidence C by $\mathbf{F_\tau=inform(A,B:C)}$.

F_τ is the reflection of an impact and we consider it as sub-entity of B.

Let denote by $\mathbf{r=evidence(A,B:C)} \in R_C$ the state of entity $C=(E_C,R_C)$ in which there exist the evidence F_τ=inform(A,B:C).

In [Markov et al., 2003a] we have formally defined the concept "interactive reflection".

Definition 4. If V_{AB} is an interactive reflection of between entities A and B, and entity C contains reflection evidences for all reflections of V_{AB} than C is called **information witness**.

Activity

Every forming relationship as well as every relationship unites the entities and this way it satisfies some theirs possibilities for building the relationship by establishing the contact. In other words, for creating the forming relationship we need:

- entities, from which the new entity is able to built;
- possibilities of the entities for establishing the contact by satisfying of which the forming relationship may be originated.

The forming relationship is the aggregate of the satisfied possibilities for establishing the contact.

It is clear that after establishing the relationship we may have any of two cases:

- all possibilities of the entities for establishing the contact are satisfied by such possibilities of other entities;
- there are any free possibilities after finishing the establishment of the new relationship - on the low levels of the entity or, if it is a new entity, on the level of the whole entity. Disintegration of the entity may generate any possibilities too.

In the second case, the entity has ***"free valency"*** which needs to be satisfied by corresponded contacts with other entities. We may say the entity has ***activity*** generated by the free possibilities for establishing the contacts with the entities from the environment.

The process of interaction is satisfying the possibilities for contact of the entities. From point of view of the entity, the interaction may be external or internal.

During the interaction given entity may be destroyed partially or entirely and only several but not all parts of the destroyed entity may be integrated in the new entity. This means that there exist both constructive and destructive processes in the process of interaction between entities. The determination of the type of the interaction depends on the point of view of given entity. The interaction dialectically contains constructive and destructive sub-processes.

If the entity is a complex, it is possible for it to have an opportunity of self-reflection. In such case, it is able to reflect any reflection, which has been already reflected in it. In this case, because of the new internal changes (self-reflection) the entity may obtain any new ***"secondary activity"***.

The secondary activity is closely connected to the structural level of the entity, which correspond to the level of the self-reflection. This way the secondary activity may be satisfied by internal or external entity from point of view of the given entity. In other words, ***the resolving*** of the secondary activity may be ***internal or external***.

Definition 5. Let $A=(E_A,R_A)$ is an entity and $r\in R_A$ is a state of A.

- $(X,\varnothing)\in\check{r}$ where $X\in E_A$, $\varnothing\in E_A$, is called **free valency** of A in the state r;
- the set Ᵽ_r of free valences for the state $r\in R_A$ is called **activity** or **expectation for contact** of A in the state r: $\text{Ᵽ}_r = \{(X,\varnothing) \mid X\in E_A, \varnothing\in E_A, (X,\varnothing)\in\check{r}\}$ ■

During the establishment of the information relationship it is possible to be generated any secondary free activity (possibilities on the low levels of the entity or on the level of the whole entity) which needs to be satisfied by corresponded contacts with other entities.

The secondary activity in the information witness generated by the information relationship is called ***"information activity"***.

Definition 6. Let $A=(E_A,R_A)$, $B=(E_B,R_B)$ and $C=(E_C,R_C)$ are entities; $F_\tau=\mathrm{inform}(A,B:C)$ is an information for A in B and $r=\mathrm{evidence}(A,B:C)$, where $r\in R_C$ is a information evidence of $\mathrm{inform}(A,B:C)$. In such case:

- $(X,\varnothing)\in\check{r}$ where $X\in E_C$, $\varnothing\in E_C$, is called **free information valency** of C based on the $\mathrm{inform}(A,B:C)$;
- the set $P_r = \{(X,\varnothing) \mid X\in E_C, \varnothing\in E_C, (X,\varnothing)\in\check{r}\}$ of free valences of the state $r\in R_C$ is called **information activity** or **information expectation** of C based on the $\mathrm{inform}(A,B:C)$.■

INFOS

On given level of complexity of the entities a new quality becomes - the existing self-reflection and internal activity based on the main possibilities for contact of the sub-entities as well as on the new (secondary) possibilities created after internal self-reflection.

The internal activity may be resolved by:

- the internal changes which lead to partial internal disintegration of the sub-entities and theirs a posterior internal integration in the new structures;
- the external influence on the environment.

The internal changes may lead to removing of some sub-entities if they have no possibilities for integration with the others, i.e. if they have no free valences to be resolved in the process of integration.

The external influence is the most important. The impact on the entities around the entity is the way to resolve its activity. The destroying of the external entities and including the appropriate theirs parts in itself is the main means to exist and satisfy the free valences.

One special kind of activity is the information one. The secondary activity need to be resolved by relevant to the information valences corresponded (information) valences. So, not every entity may be used for resolving the secondary activity.

This way, the entity needs a special kind of (information) contacts and (information) interaction for resolving the information activity.

Definition 7. The Information Witness $C=(E_C,R_C)$, which has:

- possibility for reflection in a state $r_1 \in R_C$;
- possibility for self-reflection in a state $r_2 \in R_C$;
- (primary) activity in a state $r_3 \in R_C$;
- a state $r_4 \in R_C$ in which C has non-empty information expectation (information activity)

is called **Infos.** ■

Conclusion

What gives us the concept “INFOS”?

At the fist place, this is the common approach for investigating the natural and artificial information agents.

In other hand, this is the set of common characteristics which are basic for all entities, which we may classify to the category of the INFOS.

And, at the end, this is a common philosophical basis for understanding the information subjects.

Our main goal is to provoke the scientists to continue the research in this important area and to make the next step.

Acknowledgements

This work is partially financed by project ITHEA-XXI of the Institute of Information Theories and Applications FOI ITHEA.

Bibliography

[Markov et al, 1993] Kr. Markov, Kr. Ivanova, I. Mitov. Basic Concepts of a General Information Theory. IJ "Information Theories and Applications", FOI ITHEA, Sofia, 1993, Vol.1, No.10, pp.3-10.

[Markov et al, 2003] Kr.Markov, Kr.Ivanova, I.Mitov. General Information Theory. Basic Formulations. FOI-Commerce, Sofia, 2003.

[Markov et al, 2003a] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. The Information. IJ "Information Theories and Applications", FOI ITHEA, Sofia, 2003, Vol.10, No.1, pp.5-9.

Author information

Krassimir Markov - ITHEA - FOI Institute of Information Theories and Applications; Institute of Mathematics and Informatics, BAS; P.O. Box: 775, Sofia-1090, Bulgaria; e-mail: foi@nlcv.net

Krassimira Ivanova - Institute of Mathematics and Informatics, BAS, Acad.G.Bontchev St., bl.8, Sofia-1113, Bulgaria; e-mail: foi@nlcv.net

Ilia Mitov - ITHEA - FOI Institute of Information Theories and Applications, P.O.Box: 775, Sofia-1090, Bulgaria; e-mail: foi@nlcv.net

Evgeniya Velikova-Bandova – Sofia University, Faculty of Mathematics and Informatics. e-mail: velikova@fmi.uni-sofia.bg

A TWO LAYERED MODEL FOR EVOLVING WEB RESOURCES

A. Milani, S.Suriani

Abstract*: In this paper the key features of a two-layered model for describing the semantic of dynamical web resources are introduced.*

In the current Semantic Web proposal [Berners-Lee et al., 2001] web resources are classified into static ontologies which describes the semantic network of their inter-relationships [Kalianpur, 2001][Handschuh & Staab, 2002] and complex constraints described by logical quantified formula [Boley et al., 2001][McGuinnes & van Harmelen, 2004][McGuinnes et al., 2004], the basic idea is that software agents can use techniques of automatic reasoning in order to relate resources and to support sophisticated web application.

On the other hand, web resources are also characterized by their dynamical aspects, which are not adequately addressed by current web models.

Resources on the web are dynamical since, in the minimal case, they can appear or disappear from the web and their content is upgraded. In addition, resources can traverse different states, which characterized the resource life-cycle, each resource state corresponding to different possible uses of the resource. Finally most resources are timed, i.e. they information they provide make sense only if contextualised with respect to time, and their validity and accuracy is greatly bounded by time.

Temporal projection and deduction based on dynamical and time constraints of the resources can be made and exploited by software agents [Hendler, 2001] in order to make previsions about the availability and the state of a resource, for deciding when consulting the resource itself or in order to deliberately induce a resource state change for reaching some agent goal, such as in the automated planning framework [Fikes & Nilsson, 1971][Bacchus & Kabanza,1998].

Keywords*: Temporal Resources, Dynamic Web, evolutionary resources*

Introduction

The basic notion of resource in the semantic web [Berners-Lee et al., 2001] is characterised by a unity of structure, content, and location, i.e. a resource has a structure, which is defined in the ontology, a content, i.e. the actual values of their properties, and a unique location, i.e. an URI [Berners-Lee & Fielding,1998], which uniquely identified it in term of its web location.

In our model a resource is still representing a single entity, but entities can evolve over time with respect to the current value of their contents, and also in their structural and semantic description, in other words, the notion of a resource can be intuitively intended as *the invariant aspects with respect to time of a given URI*. For example our department web page is the same resource, despite of the fact that it is continuously updated, in the content and in the structure since our web server was established in 1995.

Resources can be dated and resources can be updated. For many type of resources it is possible to specify when the information will be update, moreover the resource timestamp also provide a relevant information about its validity.

Consider for example:

a) a web page about the history of the independence war,
b) a personal CV,
c) the news of an online newspaper, and
d) stock exchange prices,

they are all web entities with a different rate of upgrade.

The advantages in explicitly defining the date/update features of a resource are apparent with respect to the trust/validity of the information provided by the resource.

Moreover consider for example a) with respect to b), and assume that these two web resources are not updated since the two years ago. It is clear that the info in a) can be used in any moment (assumed that the source is trustworthy), since we do not expect big new facts about the Independence War, on the other hand, discovering that the personal CV was not updated since two years ago, make this information not valid, thus an hypothetical software agent looking for employee information can decide to look for another CV of the same person or to ask the person to provide an upgraded copy.

Update rate can be estimated for c) and d), online newspapers and stock exchange prices are update at different pace, in the first case the content and structure can completely change after some hours, while in the latter the actual value of the price is the only thing which is likely to change, very rapidly when the stock market is open, and to remain still until next opening, during stock market closing hours. A software agent can exploit this information for its cognitive purposes by browsing the online newspaper by the hours or by the week (e.g. sport news about football matches) and the stock prices by the minutes.

Resource States

The state of a resource is an abstract characterisation of structural properties and actual values, which significantly characterise the resource. Associated with the state, there is the possibility that the resource evolves over time by moving from one state to another, in a transition path, which describes the dynamic evolution of the resource.

In the first instance a resource state is an ontological category which is simply characterised by logical constraints about the values of structural properties provided by the ontology, i.e. different ontological concept which share the same schema but not the same actual values. For example FatMan and SlimMan are instances of the concept of Man, they can be defined by constraints over the values of the properties Weight and Height, the interesting aspect is that the FatMan and SlimMan has a dynamical relationship since an individual can move from one state of another by upgrading its weight, (and less probably its height).

In the most general case, resources can allow structural properties to change, i.e. while moving from one state to another the resource evolve its schema.

It is straightforward to represent the admissible states, and the admissible state transitions of a given resource by a labelled transition diagram in which the label represent conditions or event over web resources or time which trigger the state transition.

Def. LTD for Web resources. A labelled transition diagram for a web resource it is a pair $\{N, \delta\}$

Where N is a set of nodes representing the states of the web resource, and δ represents the labelled arcs of the diagram, i.e. the state transition function which defines for every pair of nodes $n_1,n_2 \in N$ a condition L, which labels the arc (n_1,n_2), condition L is a condition over web resources (static and dynamic items, operations, web services, conditions over property values etc.) and time conditions (i.e. current date or general date/time functions).

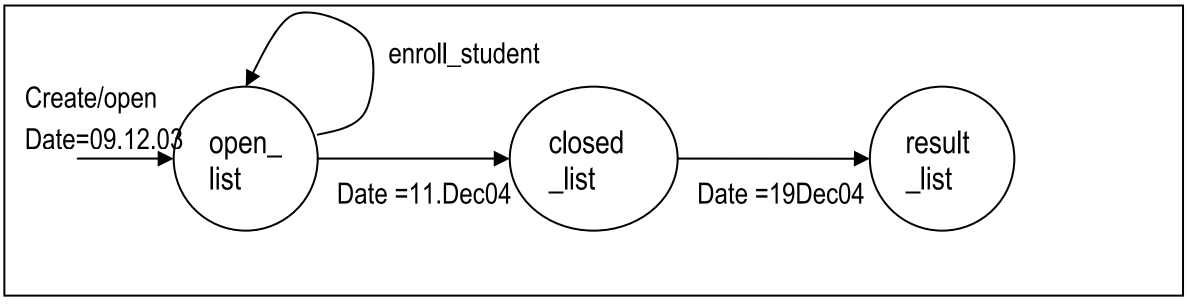


Fig.1 Students List State Transition Diagram

The labelled arc (L, n_1,n_2) denotes the fact that a resource can move from state n_1 to state n_2 when condition L is met, i.e. when the guard is verified, the logical conditions are true or the specified events take place.

By convention (false, n_1,n_2) denotes the fact that no transition is possible between state n_1 and state n_2.

A self-reference loop will represent a resource update; i.e. the resource is remaining in the same state while possibly changing its informative content.

Consider for example the online student enrolment list, exam_list, for the exam code 503 Programming Languages Course which will be held on 12th December 2004, this web resource it is continuously updated since its opening time 3 days before the exam and it is closed the day before and finally it is updated one more time with the list of candidates grades one week after the exam.

The exam_list it is an individual entity despite of the upgrade operations, which are operated on it.

In term of state transitions the evolution of the list can be represented by the state transition network in the figure 1. It is worth noticing that the self-reference loop labelled enroll_student represent the fact that after an enroll_student event.

In this framework the dynamics of resources are represented by appropriate state transition diagrams, which model the resource lifecycle.

A Two Layer Model for Dynamical Web

In order to give an account of the static and dynamic relationships of web resources a two-layer model architecture is proposed in which state transition diagrams are defined over a given ontological network.

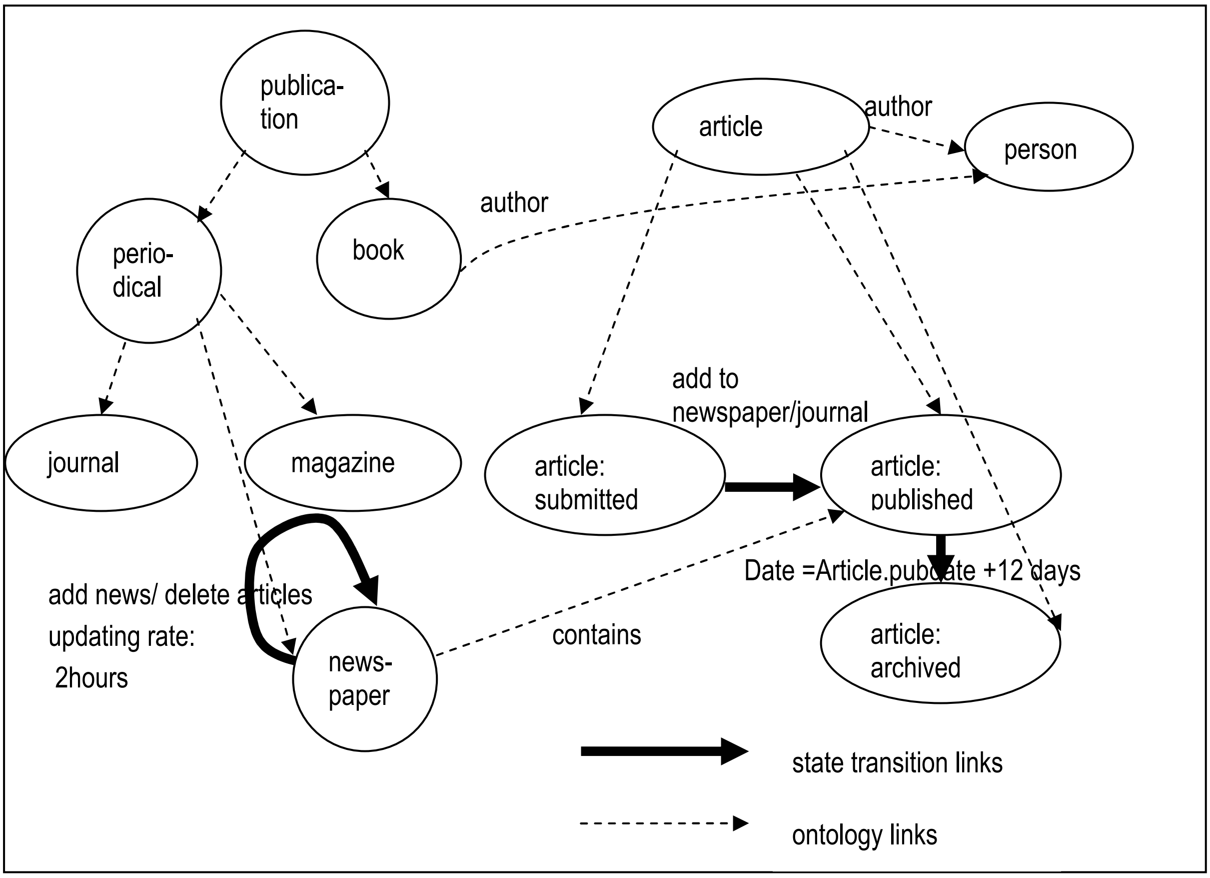


Fig.2 Students List, newpaper and articles state diagrams

For example in figure 2 is represented an online newspaper represented where news change, say, every two hours. In this case the continuous updating of the newspaper resource is represented by a single self reference labelled by operation, i.e. events add/delete news and the conditions that two hours are passed by the previous updated, the newspaper, is related by other ontological concepts (ontological link are represented by dashed lines) such as relationships of type *subclass* with respect to *periodical* and relation *contains* with respect to *published article*.

Articles inherit general concepts such as *author* relationship and can also have a more complex life cycle expressed by a LTD (submitted, published, archived) partially independent from newspaper.

Resource state transitions can be enabled by conditions over other web resources or not web resources.

Reasoning and Acting on Dynamic Resources

State transition diagrams and labelled transition diagrams are formalism popular in the area of process modelling and concurrent system modelling [Gogolla & Parisi-Presicce,1998][Cardell-Oliver et al.,1992].

The labelled transition diagram of a given resource can also be seen as an equivalent deterministic finite state automata DFA, whose transition guards (i.e. the logical conditions) denotes set of symbols in the alphabet defined by the possible binary combination of all atomic conditions. A given guard denotes the set of symbols, which correspond to atomic conditions, which makes the guard itself true.

It would be interesting to investigate the possibility of applying techniques of linear temporal logics [Manna & Pnueli,1991][Vardi,1991] used in circuit testing in order to evaluate LTL queries over a particular state of the dynamical resources.

For example an agent which has found the resource can reason about the truth of LTL modal formula such as Possibly(S) where S is a desired state of the resource, taking appropriate measure, such as *abandoning* the resource if the desired state is unreachable (e.g. the conference submitting deadline is over then transition to state "submitted" is impossible), *waiting* if the state transition is a matter of time or of exogenous events (e.g. wait until tomorrow for the President elections results), inducing the state transition by agent *deliberative action* [Milani & Ghallab,1991](e.g. reserve a ticket in order to buy it), *maintain conditions* which avoid unlikely transitions, consult the resource if the update rate or the type of time validity requires it (i.e. refresh the stock exchange prices, check again the weather forecast service).

Conclusion

The presented preliminary model extends the ontology-based approach to the semantic web, in order to represent the dynamical aspects of web resources, which evolves over time. The classical semantic web hypothesis of web resources as identified by URI is no more valid when web resources evolve; i.e. they assume different states. States of resources are represented by ontological concepts, while labeled transition diagrams are used for description admissible states of resources. The transitions are labeled by the conditions which trigger the state transition, i.e. operation, events or conditions over other resources. The resulting semantic description of web resources consist of a two layered graph which represents both static (i.e. the ontological concepts) and dynamic (i.e. the labeled transitions) relationships among concepts.

The finite state machine model allows to employ powerful reasoning technique in the semantic web graph, as for example LTL (linear temporal logic) in order to make prevision about state of web resources over time. Moreover, this rich knowledge description network can be exploited by web agents who use planning techniques for triggering the desired state transitions.

Bibliography

[Bacchus & Kabanza,1998] Bacchus,F. Kabanza,F., 1998 Planning for temporal extended goals, Annals of Mathematics and Artificial Intelligence, 22:5-27, 1998

[Berners-Lee & Fielding,1998] T. Berners-Lee, R.Fielding, U.C. Irvine, L.Masinter, Uniform Resource Identifiers (URI): Generic Syntax, Request for Comments: 2396, the Internet Society, August 1998

[Berners-Lee et al., 2001] L.Tim Berners Lee, J. Hendler, O. Lassila: The Semantic Web, Scientific American, May 2001

[Boley et al., 2001] Harold Boley, Said Tabet, and Gerd Wagner. Design Rationale of RuleML: A Markup Language for Semantic Web Rules. In International Semantic Web Working Symposium (SWWS), 2001.

[Cardell-Oliver et al.,1992] Rachel Cardell-Oliver, Roger Hale, and John Herbert. An embedding of timed transition systems in HOL. In Higher Order Logic Theorem Proving and its Applications, pages 263--278, Leuven, Belgium, Sept 1992

[Fikes & Nilsson, 1971] R.E.Fikes, N.J.Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving, Artificial Intelligence, 2(3/4), 1971

[Gogolla & Parisi-Presicce,1998] Gogolla, M. and Parisi-Presicce, F., 1998, "State Diagrams in UML - A Formal Semantics using Graph Transformation", Proceedings ICSE'98 Workshop on Precise Semantics of Modeling Techniques (PSMT'98),

[Handschuh & Staab, 2002] S. Handschuh and S. Staab. Authoring and annotation of web pages in CREAM. In The Eleventh International World Wide Web Conference (WWW2002), Honolulu, Hawaii, USA 7-11 May, 2002

[Hendler, 2001] J. Hendler, Agents and the Semantic Web, IEEE Intelligent Systems, vol.16, no.2, Mar./Apr. 2001, pp.30-37.

[Kalianpur, 2001] SMORE - Semantic Markup, Ontology, and RDF, Editor Aditya Kalyanpur A. V. Williams Building College Park, Maryland 20742

[Manna & Pnueli,1991] Manna, Z., Pnueli, A. The temporal logic of reactive and concurrent systems: Specification. Springer Verlag, 1991

[McGuinnes & van Harmelen, 2004] D.L. McGuinness F.van Harmelen, OWL Web Ontology Language Overview W3C Recommendation 10 February 2004

[McGuinnes et al., 2004] D.L. McGuinness, R. Fikes, J. Hendler, and L.A. Stein. ``DAML+OIL: An Ontology Language for the Semantic Web''. In IEEE Intelligent Systems, Vol. 17, No. 5, pages 72-80, September/October 2002

[Milani & Ghallab,1991] A.Milani, M.Ghallab eds. "New Direction in AI Planning", IOS Press 1996

[Vardi,1991] M. Vardi. An automata-theoretic approach to linear temporal logic. In F. Moller and G. Birtwistle, editors, Logics for Concurrency, pages 238-266. Springer Verlag, 1996

Author information

Alfredo Milani - Department of Mathematics and Informatics, University of Perugia, Via Vanvitelli, 1, 06100 Perugia, Italy; e-mail: suriani@dipmat.unipg.it

Silvia Suriani - Department of Mathematics and Informatics, University of Perugia, Via Vanvitelli, 1, 06100 Perugia, Italy; e-mail: suriani@dipmat.unipg.it

XML EDITORS OVERVIEW AND THE CHALLENGE TO BUILD XML-ORIENTED EDITOR FOR MEDIAEVAL MANUSCRIPT DESCRIPTIONS[14]

P. Pavlov

Abstract: *The paper presents an overview of XML and software tools for its use, with an emphasis on XML editors. Based on the experience of two Bulgarian projects on preparing electronic descriptions of mediaeval manuscripts from the 1990es, we define the following requirements to the editor used for manuscript cataloguing: minimum elements on the screen; arrangement of elements according to the practice in the subject domain; supplying default values whenever this is possible; supplying possible values in combo boxes whenever this is possible; and ease of data entry (in Bulgarian with possibility to enter mediaeval text fragments in Old Cyrillic). These requirements were taken into account for the development of a specialized editor, XEditMan, which is presented in the article. Currently, 200 descriptions of manuscripts are available which were entered and edited using XEditMan. The average time for data entry with the editor is about three times less than the time spent in previously used software tools in Bulgaria.*

Keywords: *XML, XML editors, mediaeval manuscript cataloguing, XEditMan.*

Introduction

The interest to digitisation of scientific and cultural heritage has been considerably growing in the last decades. The electronic access to cultural heritage is one of the priority areas of the European Commission. The Cultural Heritage Applications Unit of the Information Society Directorate General of the European Commission promoted the priorities in the field through a document known as *The Lund Principles* which put emphasis on *making visible and accessible the digitised cultural and scientific heritage of Europe*; *coordination of efforts*; *development of a European view on policies and programmes*, as well as of *mechanisms to promote good practice in a consistent manner* [Lund Principles, 2001]. Currently, most large institutions from the cultural sector are taking measures to make their collections available online. The first step in this direction is to provide access to cataloguing information about the holdings in a specific collection. In Bulgaria, this still is not done for the manuscript collections of any repository.

One of the recognised approaches on world wide scale is to use XML to present data on manuscripts. In this paper, we first give a brief overview on XML and tools, which allow its use. Then we present the experience of two Bulgarian projects in the field of manuscript cataloguing and formulate several basic requirements to a specialised editor for entering data on mediaeval Slavonic manuscripts and present our work in this direction.

These requirements were taken into account for the development of XEditMan, an XML editor for mediaeval manuscripts. The use of the editor is illustrated. One basic advantage is higher accuracy of entered data and better time characteristics (about three times faster data input compared to previously used tools).

Overview: XML and Various Types of Tools

XML (eXtensible Markup Language) is an open standard developed by the W3C (World Wide Web Consortium). It has two interconnected applications: web presentation and data exchange. One distinguished feature of XML is that it separates the encoding of data from program logic and user interface code. This leads to platform independence and reusability of resources.

XML is based on the Standard Generalized Markup Language (SGML), an ISO standard which puts the basics of many developments in the field of electronic transmission of documents through defining tag sets forming the DTD (document type definition) which are used to mark-up the electronic text and allow easy processing [ISO, 1986]. SGML was designed in 1986 and was oriented towards work with large collections of documents, not

[14] The research presented here is partially supported by the project KT-DigiCult-Bg (FP6) and by the ICT Agency in Bulgaria.

towards the Web. The DTD practice of SGML was expanded in XML in order to offer more data types and allow easy building of hyperdocuments. HTML was another (earlier than XML) successor of SGML designed for visualization of documents on the Web, but its orientation to present the document layout leads to limitations on the presentation of data for processing, not just for display.

An XML application unifies several parts stored separately: data, structure, presentation and program access.

The XML data are stored as the actual XML document (it is usually called the document instance). This document contains data and mark-up elements.

The second element is a file, which contains the rules defining the specific XML document's elements, attributes, entities, and processing instructions. In the beginning, following the SGML principles, a DTD file served this purpose. Later XML Schema specification started to be used in order to solve several shortcomings of the DTD: it is too restrictive to introduce new elements and does not offer support for a number of data types. XML Schema allows creating both simple and complex data types and associating them with new element types. Thus specialists working in various fields and preparing specific documents may define the structure of their documents with a great freedom.

XSL (Extensible Stylesheet Language) is the part, which ensures presentation. It allows one to render XML elements using formatting objects. For example, CSS (Cascading Style Sheets) outputs documents in HTML format. XSLT (XSL Transformation), outputs XML document into text, HTML, or any mark-up language.

The last component is called DOM (Document Object Model) which allows accessing data encoded in XML through programs. Thus data can be separated from the specific platform.

There are several types of XML-oriented tools. The *XML editors* are used to create structured documents. From the point of view of the user, it is important to have an easy and understandable interface for entering data. The task of collecting manuscript descriptions in XML format inevitably raises the question how the data will be entered. Other types of tools are necessary basically for the IT staff, such as software for the *creation of DTD's or XML Schemas*, parsers for *validating XML* files (applications which check the documents against the DTD or the Schema); parsers for *parsing XSLT* (they prepare XML documents for presentation as text, HTML or PDF by applying the XSLT stylesheet language). Technical staff may also need a specialised editor for speeding the *creation of XSLT stylesheets*. For the work on manuscript catalogue descriptions, most important are the editor and the parser. We provide below some explanation and examples of tools from the various categories.

Editors

Editors allow users to create and edit XML documents using the proper DTD. XML editors often provide additional functionality, for example validation of the document against the DTD or schema. To facilitate the user in his/her work, editors rely on two basic methods:

- Use of colours to distinguish elements, attributes, and text, etc. for easy reading.
- Providing clickable lists of possible elements and attributes at the current cursor point in the document. These lists usually are located in the left pane of the editor window.

Popular professional XML editors are XMetaL® 4[15], xmlspy® 2004[16], NoteTab Pro[17]. Free editors are XMLCooktop[18], Bonfire Studio 1.4[19], NoteTabLight[20], Xeena[21], Xerlin[22] etc. The illustration on Fig. 1. shows a snapshot from xmlspy® 2004.

Validating Parsers

Usually, the professional XML editors contain a built-in validator. Some are internal to the editor and others use a separate piece of software.

---

[15] http://www.sq.com/, last visited on 25 April 2004.
[16] http://www.xmlspy.com, last visited on 25 April 2004.
[17] http://www.notetab.com, last visited on 25 April 2004.
[18] http://www.xmlcooktop.com/, last visited on 25 April 2004.
[19] http://www.nzworks.com/bonfire/download.asp, last visited on 25 April 2004.
[20] http://www.notetab.com, last visited on 25 April 2004.
[21] http://www.alphaworks.ibm.com/tech/xeena, last visited on 25 April 2004.
[22] http://www.xerlin.org/, last visited on 25 April 2004.

XSLT Parsers

The XSLT parsers play the role of formatting engines. They output data most often in HTML, text, PDF. They are sometimes part of the XML editor.

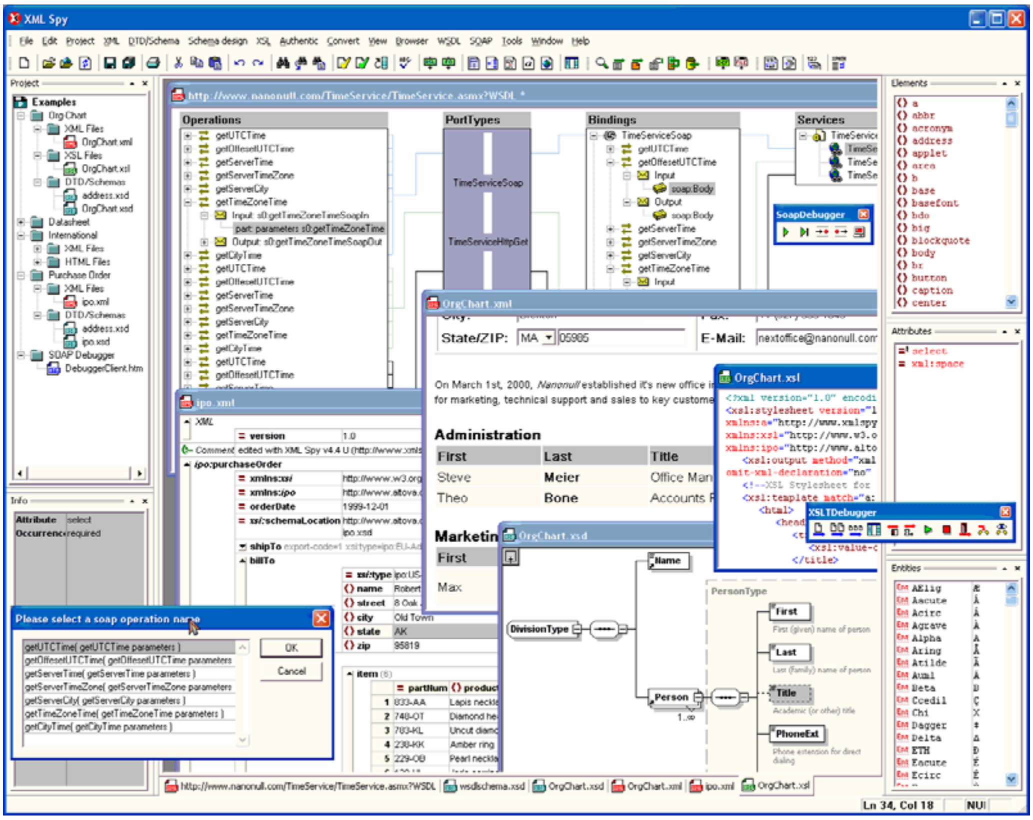


Fig. 1. Snapshot from *xmlspy® 2004, source:* http://www.xmlspy.com

The Bulgarian Experience in Preparing XML Editor for Mediaeval Manuscripts Descriptions

The idea to use a markup language for manuscript descriptions goes back to the 1990es. With the advent of mark-up languages, a team in Bulgaria suggested in 1994-95 a structured description of manuscript data built as an extension of Text Encoding Initiative [TEI] of that time. A project called *The Repertorium of Old Bulgarian Literature and Letters* was started as "…an archival repository capable of encoding and preserving in SGML (and, subsequently, XML) format of archeographic, palæographic, codicological, textological, and literary-historical data concerning original and translated medieval texts represented in Balkan Cyrillic manuscripts" [Repertorium], [Miltenova et al., 2000]. This is a typical repository project aimed to answer researchers' (not librarians') needs. The computer model based on SGML is discussed in [Dobreva, 2000]. Currently there are 300 manuscript descriptions, which should be made available on the project website[23].

In the late 90es, the National Library "St. Cyril and St. Methodius" and the Institute of Mathematics and Informatics became associated members of the MASTER project *(Manuscript Access through Standards for Electronic Records)* supported by the EC [MASTER]. Within this project, a TEI-conformant DTD for mediæval manuscripts was developed with the ambition to answer the needs of all repositories in Europe, and software for making and visualising records on manuscripts. The MASTER standard (may be with small revisions) was adopted by the TEI in May 2003.

In the Repertorium project, data were entered through Author/Editor software product of SoftQuad Company, a predecessor of HoTMetaL and currently available XMetaL editors. In the data entry process, users were seeing

---

[23] On April 25, 2004 there was still a message that link is disabled for file update.

all elements from the description on the screen (surrounded by the SGML delimiters, e.g. <P> </P>) which formed long list spread on several screens. This was not very convenient, if we also add that the appearance of elements followed the structure of the DTD, which is not the same as the sequence of elements natural for the people working with mediaeval manuscripts. The organization of work was oriented towards one specialist working on one description, which produced results of different quality in the group of almost 10 specialists working on the descriptions [Dobreva, Jordanova, 2000]. The description data were entered in English which made them usable by English language speakers. To enter fragments of Old and Middle Bulgarian texts a designated font was created, and in data entry the LANG attribute was assigned to elements containing text in Old or Middle Bulgarian while for all other languages was supposed that they contain texts in English.

The experience of the pilot catalogue descriptions within the MASTER project was different in two directions: the data were entered in both Bulgarian and English with the idea that this will serve larger research community, and the editor used for the tests was NoteTabLight[24] (see Fig. 2). To enter data on both languages, elements were repeated with including of the LANG attribute showing the language of the data entered within the specific element.

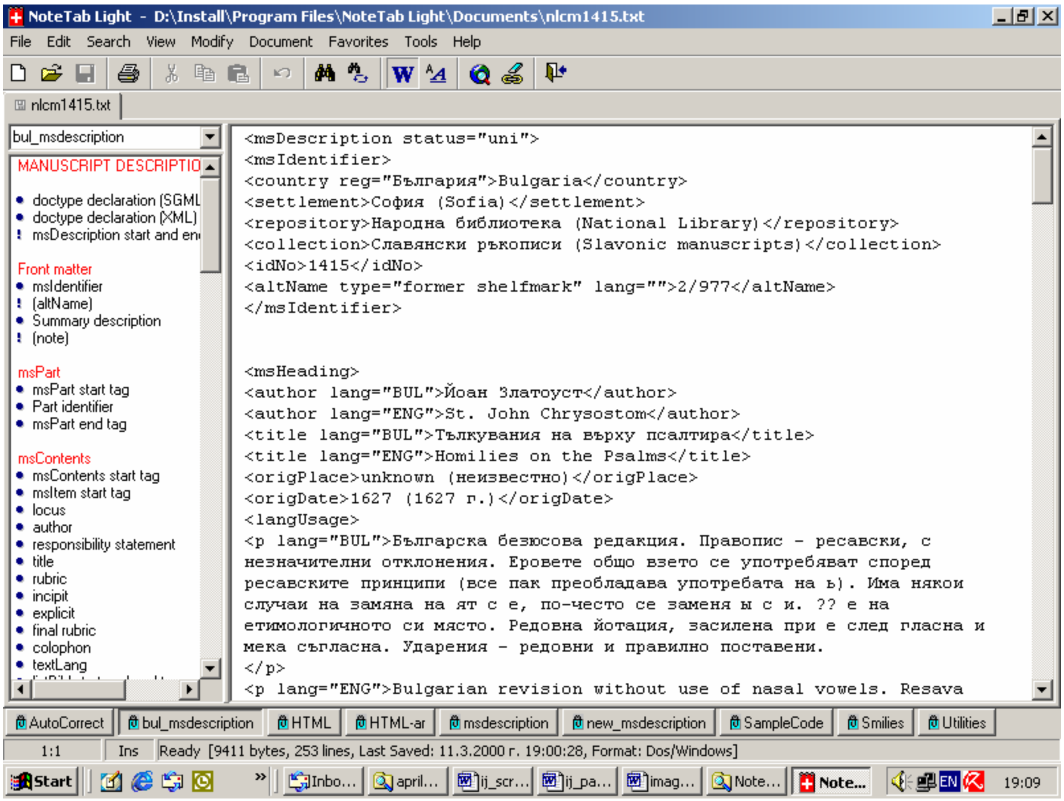


Fig. 2. Example of data entry of manuscript descriptions in NoteTabLight

Unlike the Author/Editor interface, in this case all available elements can be seen on the left pane of the window, and the person who enters data should click on the specific element which is needed. This led to high number of erroneously located elements and a heavy workload on editing the descriptions.

The experience from both projects stimulated us to formulate the following requirements to a specialized editor:

- The number of elements visible on the screen should be the minimum possible. Lengthy lists of elements confuse users who are specialists in mediaeval studies or library cataloguing who normally would enter the data. This also slows down the process of data entry and leads to mistakes.
- The sequence of appearance of the elements should follow the logic of the subject domain, not of the XML DTD.

[24] http://www.notetab.com/ntl.php, NoteTabLight – free editor offered as an alternative to the commercial professional editor NoteTab Pro.

- o Quite often, the value of element is “No information” (this is because in some cases there is no information on the matter since these descriptions are based on preliminary research work on the manuscripts). To avoid multiple entry of this phrase, the value can be supplied in advance and changed by the person who enters data whenever this is needed.
- o There are several elements where the values are chosen from a list: for example, names of repositories, cities, values of attributes for language, etc. To avoid errors, combo boxes with possible values could be supplied.
- o Ease of entering data written in Old Cyrillic script.
- o Interface in modern Bulgarian (thus specialists who enter the data see names of elements which are familiar to them, and do not have to become acquainted in details with the DTD itself.

Taking these considerations into account, we decided to create a specialized editor, which takes into account these requirements in its interface. The decision to create a home-made editor was taken after the consideration of possibilities to adapt existing commercial editors. Since the left pane with all elements listed and the alphabet encoding could not be solved satisfactorily, we decided to create a tool which could be easily installed on a computer with a running Microsoft Internet Explorer browser and Internet Information Server.

XEditMan: A XML Editor for Mediaeval Manuscripts Descriptions

XEditMan is actually a set of tools: editor for new document, editor for existing document and a visualisator. The editor is currently oriented towards the use of the MASTER DTD for manuscript descriptions adopted by TEI[25].

Data Entry: The New Document Editor and the Editor for an Existing Document

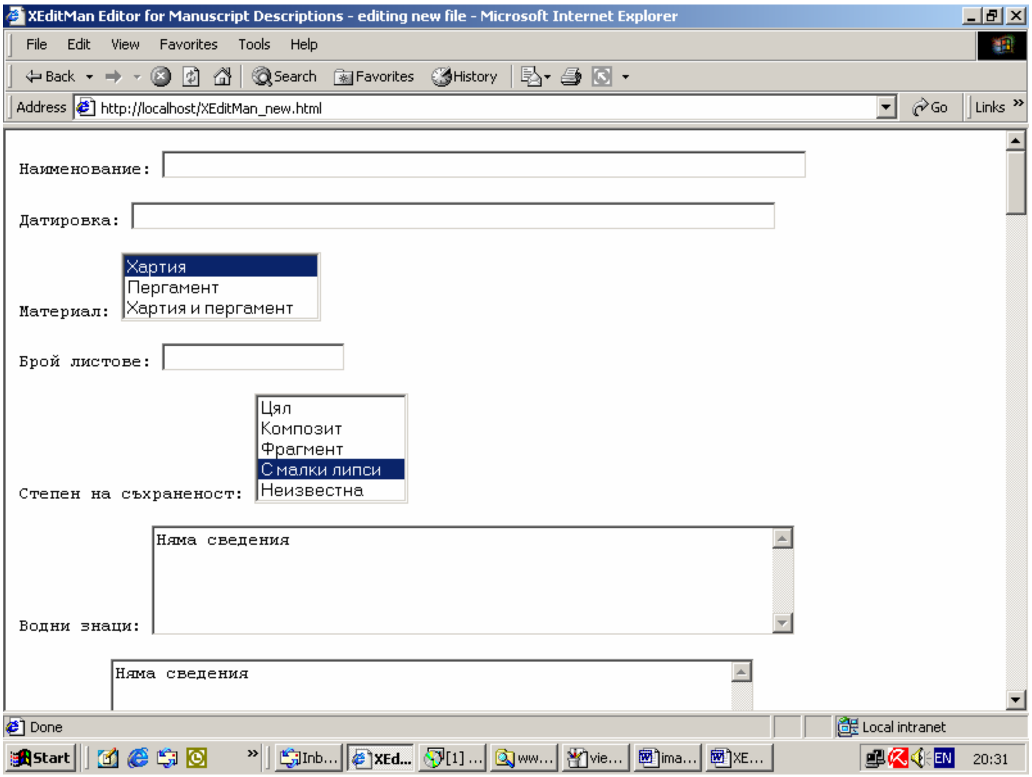


Fig. 3. XEditMan: Data entry interface

The editor of new document is used to enter data arranged in the order, which is natural for the subject domain. In two cases repetitive elements are possible: description of scribes and description of texts appearing in the

[25] The relevant materials can be found on http://www.tei-c.org.uk/Master/Reference/, last accessed on April 25, 2004.

manuscript. In these cases, during the first entry the user supplies the data on the first scribe (respectively, text) and the total number of scribes (texts). Then when the description is opened with the editor of existing texts, the respective number of elements appear in the window and make possible the entry of the information on the other scribes (respectively texts). Fig. 3 presents part of the data entry window, in which we see several types of elements: with no value; with supplied values, and combo boxes for choice of possible value.

The first two fields on Fig. 3, name and date, are typical fields for direct data entry. The third and fifth elements, material and manuscript status, are supplied with combo boxes containing possible values. In the last two elements (the visible one is a watermark) the value "No information" is entered by default. If there is no information about the element, the specialist who enters data does not have to bother with writing this text again and again.

After the data are entered, the users clicks a button "Save the description" which generates the XML document conformant to the MASTER project DTD (see Fig. 4); now all element identifiers appear according to the DTD.

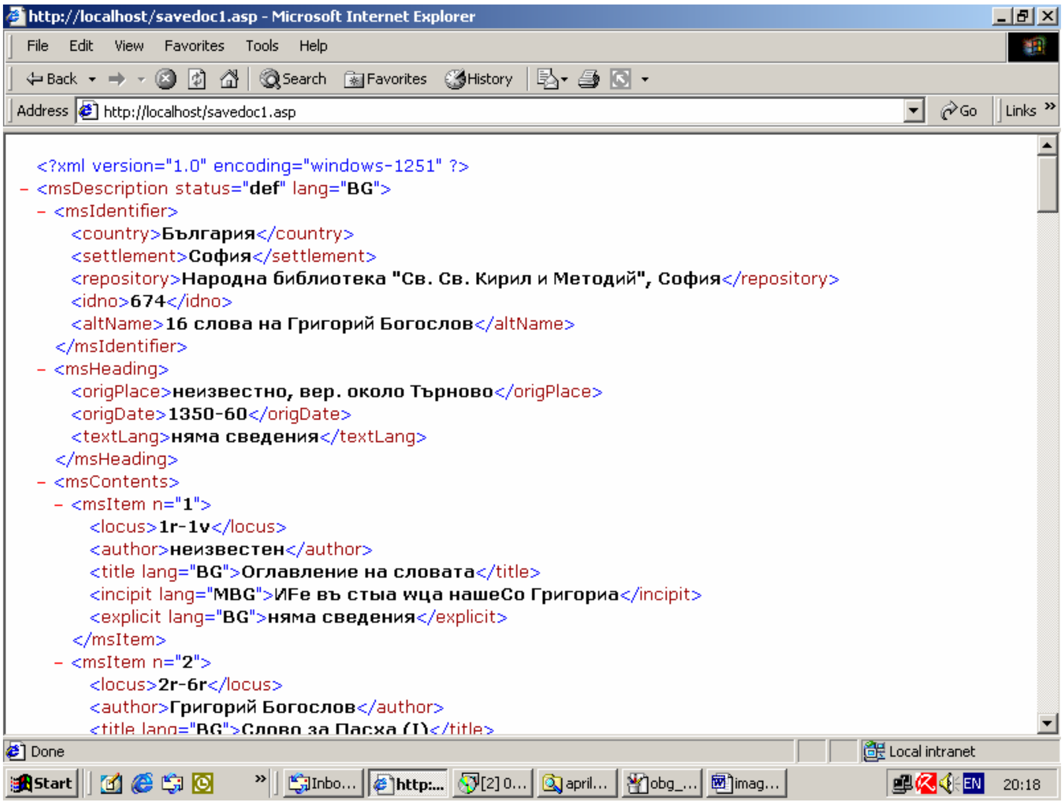


Fig. 4. A Sample of XML Document which is being saved after data were entered in XEditMan

The generation of this document is done in a way, which guarantees successful validation. This organization of work combines easy data entry and DTD-conformant result. For this reason, the editor does not include an internal validator. It is suggested to use a commercial editor for validation purposes and for cases where the interface of the editor does not support too specialized cases appearing sometimes in manuscript descriptions, like quoting within the content of specific element. We made experiments with the use in such cases of TurboXML editor (see Fig. 5). The work on XEditMan was done with the idea to cover the mass case of data entry on manuscripts. In very specific cases which appear rarely (like nesting quotes, bibliographical references and corrections to the Old Bulgarian texts), but would require too many complications in the interface, specialists who are familiar with the DTD could enter data using commercial editors which arrange the document as it is saved in XML format.

Data Visualisation

After the data are entered, they can be visualized with the help of another component of the editor. There are two modes of visualization: visualization of the complete document, which is demonstrated on Fig. 6, or visualization of selected elements from the description.

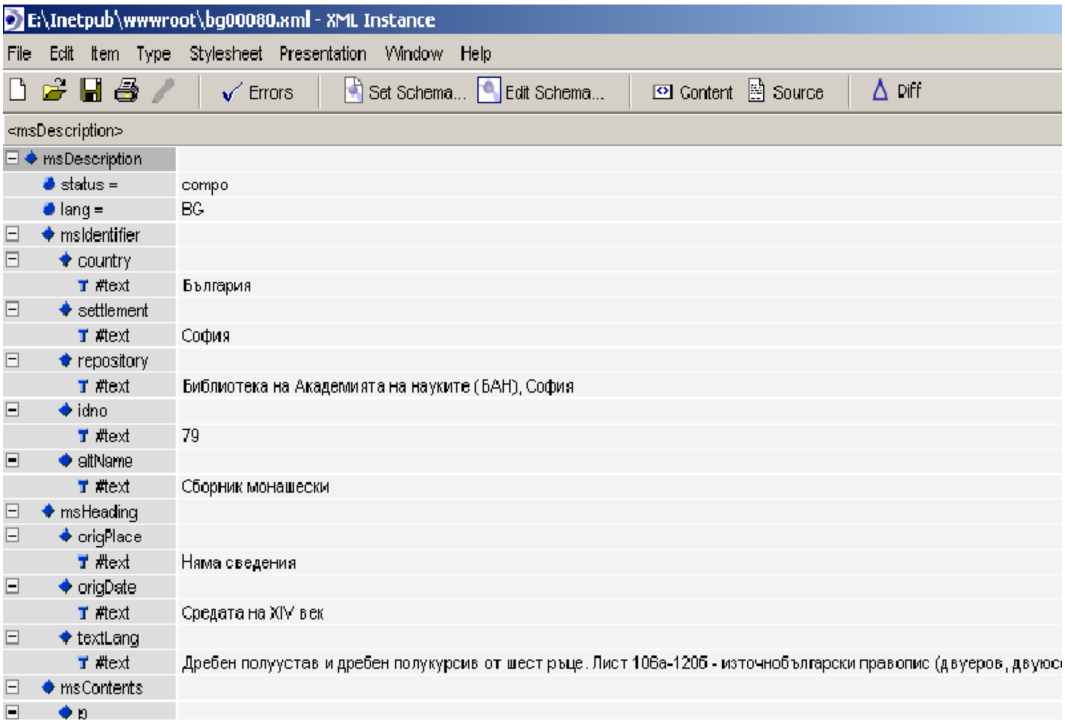


Fig. 5. An Example of Description Prepared in XEditMan Visualised in TurboXML editor

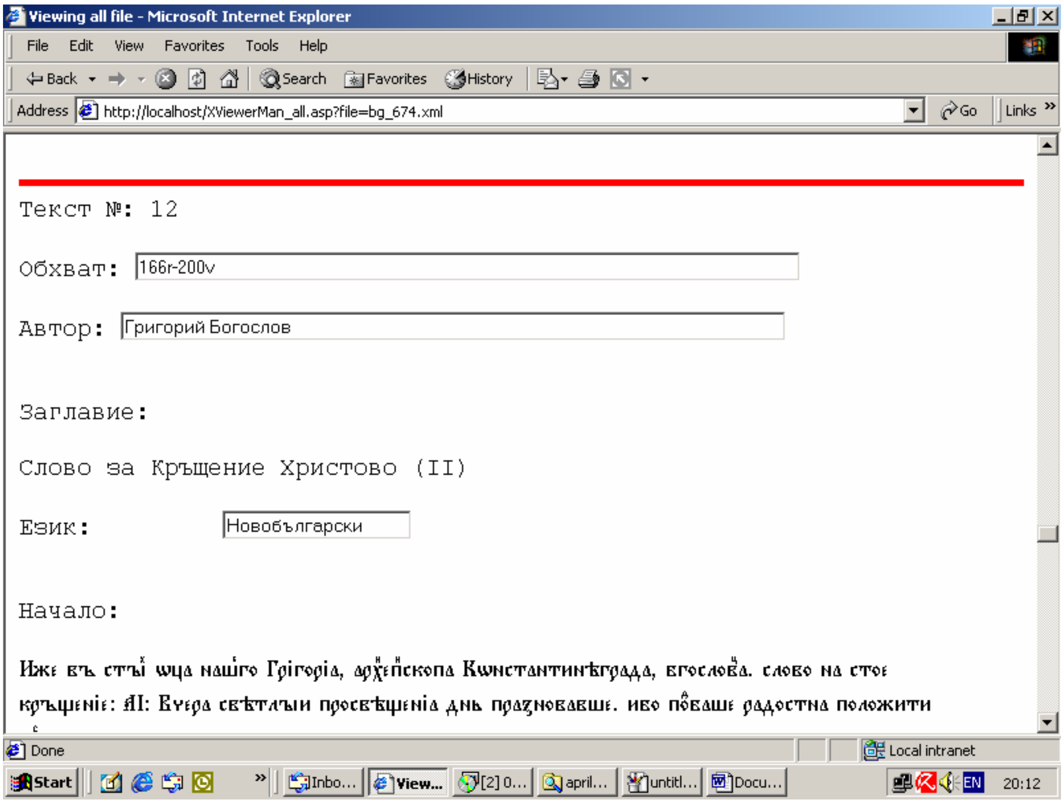


Fig. 6. An example of Visualized Data

To make possible further processing of data on sets of manuscript descriptions, we are currently working on a program interface, which would extract data from XML descriptions into a database. This would provide tools for group queries.

Conclusion

The paper presented a brief overview of XML and the current trends in developing tools for its use. It formulated several basic requirements for the development of a specialized editor on mediaeval manuscripts, which guarantee faster and more accurate data entry.

It also presented the experience of the author in designing XEditMan, a specialized editor for manuscript descriptions. XEditMan was tested in the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences and is now used to enter data on Bulgarian manuscripts stored in Bulgaria. Two hundred descriptions are already available by the date of preparation of the paper (25 April 2004) based on the catalogue [Ikonomova et al., 1982].

This work is made as part of the current project Knowledge Transfer for the Digitization of Cultural and Scientific Heritage to Bulgaria, coordinated by the Institute of Mathematics and Informatics and supported by the Framework Programme 6 of the European Commission.

The basic idea is to provide in the next months a set of 800 manuscript descriptions which form about 1/10 of the manuscripts stored in Bulgaria. The first group of manuscripts, which was chosen, consists of Bulgarian manuscripts.

This work is extensible in two ways – more manuscripts could be added to the collection, and more data could be supplied at a later stage. For this reason, we believe that this initiative will contribute to the more adequate presentation of the cultural heritage of Bulgaria.

Bibliography

[Dobreva, 2000] M. Dobreva, A Repertory of the Old Bulgarian Literature: Problems Concerning the Design and Use of a Computer Supported Model, In: A. Miltenova, D. Birnbaum (eds.), Medieval Slavic Manuscripts and SGML: Problems and Perspectives, Sofia, Academic Publishing House, 2000, pp.. 91-98.

[Dobreva, Jordanova, 2000] M. Dobreva, M. Jordanova, *Some Psychological Aspects of Computer Modeling of Complex Objects,* In: A. Miltenova, D. Birnbaum (eds.), Medieval Slavic Manuscripts and SGML, Problems and Perspectives. Prof. M. Drinov Academic Publishing House, Sofia, 2000, pp. 295–310.

[Ikonomova et al., 1982] A. Ikonomova, D. Karadzhova, B. Christova, Bulgarian Manuscripts from 11 to 18 cent., stores in Bulgaria. Vol. 1. Sofia, 1982. (In Bulgarian – Икономова, А., Д. Караджова, Б. Христова. Български ръкописи от XI до XVIII век, запазени в България. Своден каталог, том I, НБКМ, София, 1982.)

[ISO, 1986] International Organization for Standardization, *ISO 8879: Information processing – Text and office systems - Standard Generalized Markup Language (SGML),* Geneva, ISO, 1986.

[Lund Principles, 2001] http://www.cordis.lu/ist/ka3/digicult/lund_principles.htm—*eEurope: creating cooperation for digitisation (Lund Principles)*

[TEI] http://www.tei-c.org/—*Text Encoding Initiative Website*

[MASTER] MASTER, http://www.cta.dmu.ac.uk/projects/master/, website of the MASTER project.

[Miltenova et al., 2000] A. Miltenova, D. Birnbaum (eds.), *Medieval Slavic Manuscripts and SGML: Problems and Perspectives*, Sofia, Academic Publishing House, 2000, 372 pp.

[Repertorium] *Repertorium*, http://clover.slavic.pitt.edu/~repertorium/index.html — website of the *Repertorium of Old Bulgarian Literature and Letters*

Author information

Pavel Pavlov – Sofia University, Faculty of Mathematics and Informatics, Assistant Professor; 5 J. Bouchier Blvd, Sofia, Bulgaria; e-mail: pavlovp@fmi.uni-sofia.bg

LOCAL GOALS DRIVEN HIERARCHICAL REINFORCEMENT LEARNING*

A. Pchelkin

***Abstract**: Efficient exploration is of fundamental importance for autonomous agents that learn to act. Previous approaches to exploration in reinforcement learning usually address exploration in the case when the environment is fully observable. In contrast, the current paper, like the previous paper [Pch2003], studies the case when the environment is only partially observable. One additional difficulty is considered – complex temporal dependencies. In order to overcome this additional difficulty a new hierarchical reinforcement learning algorithm is proposed. The learning algorithm exploits a very simple learning principle, similar to Q-learning, except the lookup table has one more variable – the currently selected goal. Additionally, the algorithm uses the idea of internal reward for achieving hard-to-reach states [Pch2003]. The proposed learning algorithm is experimentally investigated in partially observable maze problems where it shows a robust ability to learn a good policy.*

***Keywords**: reinforcement learning, hierarchical behaviour, efficient exploration, POMDPs, non-Markov, local goals, internal reward, subgoal learning.*

Introduction and Problem Statement

One of the directions in artificial intelligence is adaptive autonomous agents' research. This research direction started growing actively in 1985 [Maes95,Wil85], however, it was proposed to make researches in similar directions also before it [Bong75]. Reinforcement learning [Sut98,KLM96,Mit99] examines a question how an autonomous agent [Maes95] that senses and acts in its environment can learn to choose optimal actions to achieve its goals. The agent can perceive a set of distinct perceptions from its environment and has a set of actions that it can perform. At each discrete time step, the agent senses the current percept, and chooses an action to perform it. The environment responds by producing the succeeding state and the agent can perceive a new observation. If the agent achieves the goal state, the environment gives the agent a reward.

The set of actions allowed to the agent is fixed and defined before learning. The structure of the environment is unknown to the agent and is represented by a black box. This means that it has to obtain all knowledge helping to achieve the agent's goals only by itself, only by experimenting with the environment. The task of the agent is to perform sequences of actions, observe their consequences, and to learn a control policy.

Irreversible transactions: Efficient exploration plays a fundamental role [Thr92b] for autonomous agents that learn to act. In many reinforcement learning algorithms undirected exploration techniques are used. While undirected exploration techniques, *e.g.* random walk exploration, utilize no exploration-specific knowledge and ensure randomness into action selection, directed techniques rely on knowledge about the learning process itself, allowing for exploring in a more directed manner [Thr92a]. In many finite deterministic domains, any learning technique based on undirected exploration is inefficient in terms of learning time, *i.e.* learning time is expected to scale exponentially with the size of the state space [Whit91].

The reason for the difficult exploration by undirected techniques is the existence of irreversible transactions between states of the environment. Usually learning algorithms are being investigated in fully reversible domains [Sut98,KLM96,Mit99,McCallum95], *e.g.* the maze problem. In the maze problem each action has an opposite action, and the agent usually can easily undo its previous action by one step, *e.g.* the action "go left" can be undone by the action "go right". Because of this reason, each state of the environment becomes easily achievable. However, it could not be easily applied to the real case, *e.g.* if the goal of the agent is to build a house from blocks, one wrong random action may discharge all the previous work by destructing the built construct. That's why, the existence of irreversible transactions is not exclusion or the specifically invented difficulty but it is a real difficulty that was unfortunately ignored in many investigations making simulations in the artificially designed environments.

* This research was partially supported by the Latvian Science Foundation under grant No.02-86d.

Usually in practice multiple situations are indistinguishable from immediate perceptual input. These multiple situations may require different responses from the agent. Usual reinforcement learning techniques, such as Q-learning [Wat89], can't be applied in partially observable domains. Due to this reason, McCallum has developed a learning algorithm "Utile Suffix Memory" (USM) [McCallum95] that is able to overcome the incomplete perception problem in order to learn a near-to-optimal policy in partially observable environments.

Efficient exploration in partially observable domains is a special difficulty [Chr92]. Previous approaches [Thr92a,Sut90] to exploration in reinforcement learning usually address exploration in the case when the environment is fully observable. In contrast, McCallum [McCallum97] considers efficient exploration applied to USM in the partially observable environment. In our study, we continued his research by applying to USM an exploration technique based on internal reward for hard-to-reach states [Pch2003], and our modification has outperformed the original algorithm in the case of difficult exploration.

Problem statement: Unfortunately, USM has one very important drawback – it is not able to make perceptual distinction by seeing too far future back in time. That's why, USM is not able to discover too complex temporal dependencies that include too many time steps. That was a motivation for the current paper to develop a reinforcement learning algorithm having an ability to overcome the combination of all the difficulties described above, *i.e.*:

- incomplete perception;
- irreversible transactions;
- complex temporal dependencies.

Automatic building of hierarchies: Complex temporal dependencies are usually solved by allowing the agent acting hierarchically [WS98,SS2000]. That's why, the present paper also considers hierarchical reinforcement learning as a key method for overcoming difficulties related with complex temporal dependencies.

There are many different models of hierarchical reinforcement learning [KLM96]. Considering different approaches of making hierarchical reinforcement learning, it is possible to distinguish two cases: (1) the use of structurally pre-determined domain-specific hierarchies and (2) automatic building of hierarchies. Most of hierarchical reinforcement learning algorithms are based on an assumption that fixed domain-specific knowledge about hierarchies is provided and can be exploited by the algorithm, *e.g.* [Die2000,PR97,Hum96] are only a small part of them. In contrast, in the current paper it is assumed that no prior domain-specific knowledge about subtask hierarchies is provided to the agent. Like [WS98,SS2000], the current problem statement is more difficult, but it is also more realistic.

Learning algorithm

Training skills at local goals: In previous paper [Pch2003], it has been found that McCallum exploration techniques [McCallum95,McCallum97] may fail in the case when the environment isn't reversible, *e.g.* if there are one-direction ways. In the last case, it may be difficult to find a goal first time. USM uses Q-values to discover distinctions in the environment, but these Q-values are accessible only when the agent has reached the goal and has received the reward from the environment at least one time. Until this, the agent is unable to discover history distinctions and, thus, is unable to overcome incomplete perception problem. This problem has been solved [Pch2003] by giving the agent additional internal reward for state space exploration. Receiving additional internal reward for exploration USM was able to optimise its control policy not only for exploiting the environment but also for exploration in the same manner. It was relied on hypothesis that the perceptual distinctions discovered during exploration will help the agent to reach the goal state. In general, there is no principal difference between exploitation and exploration because in both cases the goal is to reach some special states of the world. In many cases distinctions needed for reaching the goal state are also needed for reaching some particular state. Simulation results in the maze domain had successfully confirmed the hypothesis.

The agent has the goal defined by its environment, let's call it the *global* goal (externally defined goal). However, any other state of the environment can also be considered as a goal, and it could be called a *local* goal (internally set goal). In the previous paper, while the agent was not able to reach the global goal, it was trying to reach some local goal to advance its skills of the environment control. Training skills at reaching local goals had helped the agent to obtain skills sufficient for achieving the global goal. In the current paper, it is proposed to use the same idea of learning to reach local goals when the global goal is not achievable or it is hardly achievable.

Key ideas: The proposed learning algorithm is based on two ideas:

- hierarchical behaviour;
- training skills at local goals.

The agent needs hierarchical behaviour in order to overcome incomplete perception and complex temporal dependencies. Training skills at local goals are needed to make efficient exploration when the environment has irreversible transactions.

The learning algorithm has two parts. The first part selects the current main goal using three possible reasons: (1) a need to achieve the global goal, (2) a need to explore rarely observed perceptions and (3) a need to train skills at hard-to-reach goals. The task of the second part is choosing actions or selecting subgoals (like the calls of subroutines) in order to reach the current main goal. The selection of some subgoal can also be considered as abstract actions. In order to make the learning algorithm more simple, it is proposed to learn the primitive action selection policy and the abstract action selection policy in the same manner, applying the same principles.

Observations from the environment are used twice: (1) as the context information for action selecting and (2) as subgoals. In this sense, any main goal or any subgoal is a normal perception that has been observed but at the current moment the agent is trying to reach the state producing this observation. If the current observation is equal to the current goal, the goal is considered to be achieved. This means that the proposed learning algorithm is driven by local internal goals and their subgoals, and each goal or subgoal is a usual observation temporally considered in such role.

The agent also has the memory about successful cases and the agent adaptation rule could be described as follows: if in the context of observation *p* the selection of action *a* helps to achieve the current goal *g* (that can be a subgoal at the higher level), then next time the probability of the selection of the same action in the same context (observation *p* and the goal *g*) must be increased. This means that the learning algorithm doesn't exploit dynamic programming ideas about the estimation of distance to the goal, but it performs only pure reinforcement of successful actions.

To sum up, it should be noted that the agent does not only learn when to select what action, but also - when to select what subgoal. It means that the agent also must learn subgoal selecting policy.

Formal description of algorithm: The agent has the fixed set of actions *A*. The set of perceptions *P* is not directly given to the agent, instead, it is maintained all time and contains perceptions observed by the agent till the current time moment. Similarly, the counter $c(p)$ – the number of times the agent has observed percept *p* is maintained in order to provide directed exploration. Additionally, it is proposed to maintain the degree of difficulty $d[g]$ (initially equal to zero) for each local goal $g \in P$ that stores the total number of all failures minus the number of all successes at achieving goal *g*. Consequently, $D = \{ p \in P \mid d[p] > 0 \}$ can be considered as a set of difficult goals.

During learning process, the agent maintains its lookup table with real values $q[p,g,a]$, initially equal to 1, for each $p,g \in P$ and $a \in A \cup P$. This table doesn't store estimates of expected future discounted external rewards, *e.g.* as in *Q*-learning, instead, the value $q[p,g,a]$ stores the sum of all the internal rewards for performing action *a* or selecting *a* as a subgoal (if *a* is a perception) in the context of observation *p* and local goal *g*. The internal reward is not obtained from the environment, but the agent internally generates it for obtaining local goals or subgoals. To prevent the recursive calling of the same subgoal, it is proposed to define a set of the currently selected goals in the stack, noted by *G*.

It is also assumed that there is only one goal state in the environment, and the agent maintains a variable *f* storing initially *null*, or the observation of the goal state if the goal state has been achieved. The goal state can be recognized obtaining a positive reward from the environment.

The lookup table $q[p,g,a]$ is used for action selecting. However, this table is too big and it needs some kind of generalization on its values. For example, it may have an empty cell for some action *a* in the context of goal *g* and perception *p*, but at the same time it may have learned values for the same action in another context, and the last information also can be exploited in selecting of action *a*. For this purpose, we can define generalized value $Q(p,g,a)$ as follows:

$$Q(p,g,a) = \alpha_1 \cdot q[p,g,a] + \alpha_2 \cdot \frac{1}{|P|} \sum_{p \in P} q[p,g,a] + \alpha_3 \cdot \frac{1}{|P|^2} \sum_{p,g \in P} q[p,g,a]$$

Other notations: *t* – the current time moment, *random* – random value in interval [0;1).

The proposed hierarchical reinforcement learning algorithm can be described as follows:

```
Main:
  G = ∅
  repeat
    ExecuteAction(RandomAction); TryToReachGoal(GetMainGoal,λ)

GetMainGoal:
  if f ≠ null and (D = ∅ or random ≥ δ)  then result = f
  else if D ≠ ∅ and random < ½ then
         result = select g∈D with probability Pr(g) = 1
       else result = select g∈P with probability Pr(g) = 1/c(g)^η

TryToReachGoal(g,s):
  G = G∪{g} ; t0 = t ;  i = 0 ; E = ∅
  while CurrentObservation ≠ g and i < τ
    i = i+1
    p[i] = CurrentObservation
    V = { x∈A∪P | s ≥ 1/Q(p[i],g,x) & x∉G∪E∪{p[i]} }
    if V = ∅ then V = A
    a[i] = select x∈V with probability Pr(x) = Q(p[i],g,x)^β
    if a[i] ∈ A then ExecuteAction(a[i]) ; r[i] = SUCCESS
    else r[i] = TryToReachGoal(a[i], s - 1/Q(p[i],g,a[i]))
    if r[i] ≠ SUCCESS then E = E∪{a[i]}
  if CurrentObservation = g then
    while r[i] = SUCCES and i > 0 do
      q[p[i],g,a[i]] = q[p[i],g,a[i]] + 1/(t-t0)^γ
      i = i-1
  G = G -{g}
  if CurrentObservation = g then d[g] = d[g]-1 else d[g] = d[g]+1
  if CurrentObservation = g then result=SUCCESS else result=FAILURE
```

Notation *result* means the resulting value of a function, and notations t_0, *i*, *E*, *p*[*i*], *a*[*i*], *r*[*i*], *V* are local variables of function "TryToReachGoal". The algorithm has a series of parameters: α_1, α_2, α_3, β, η, δ, γ, λ, τ, and it is proposed to use the following settings: α_1=1, α_2=0.01, α_3=0.0001, β=5, η=5, δ=0.7, γ=0.1, λ=3, τ=4.

Simulation results

The presented above learning algorithm has been tested using three different maze problems: maze1, maze2 and maze3 (see Fig.1). Each maze is a local perception grid world. The essence of this problem is searching for immovable goals in a maze. The agent's life consists of many trials: it is placed in a random empty cell, after which the agent has to find the goal (marked "G") searched with the least possible number of steps. Initially the agent has not any knowledge on the environment. Each trial can be considered as one problem solved by the agent. In the course of trials, the agent has to learn to quickly find this object.

The agent can move to nearest empty cells only (white or silver cells, but not black ones; some cells are specially highlighted with silver – it means that these cells have duplicated observations). Eight possible directions mean eight possible actions that the agent can execute. If the agent tries to move onto barrier, it stays at the same position. This creates many cycles in the environment, and makes the learning task more difficult. The agent can perceive only the containment of nearest eight cells. So, there are different, but perceptually identical, world states.

Additionally, the cells can contain special symbols - arrows. These are normal empty cells, except the agent can move only in the direction defined by a corresponding arrow (in other case it stays in the same position). These arrows are needed to simulate discussed above irreversible transactions between the environment states.

To simulate complex temporal dependencies, there are presented two special cells: a dour (marked “D”) and a key (marked “K”). To be able to come into the cell with a dour, the agent needs to visit the cell with a key before. After visiting the cell with a key, the agent is able to come into the cell with a dour only once. If the agent is not able to come into the dour, it stays at the same cell. The idea about the dour and the key has been taken from the paper [WS98].

Results: Figures 2,3 and 4 show the dynamics of the number (see “steps to goals”) of steps performed by the agent to reach the goal (a cell marked “G”) in each trial during the learning process. Abscise “problem number” means the index of a goal searching trial. All experiments were repeated 10 times, and there are presented only averaged results.

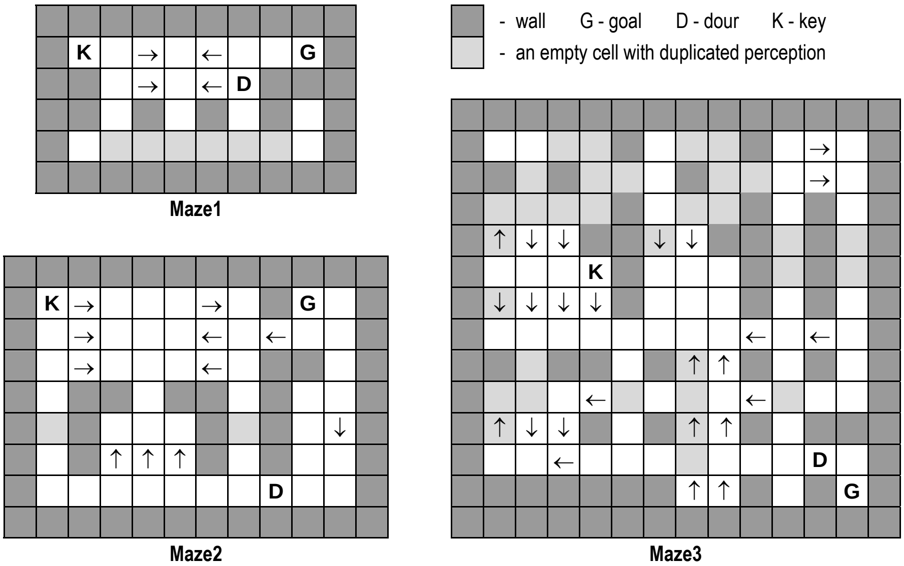


Figure 1. Three different maze problems: maze1, maze2 and maze3 for experiments

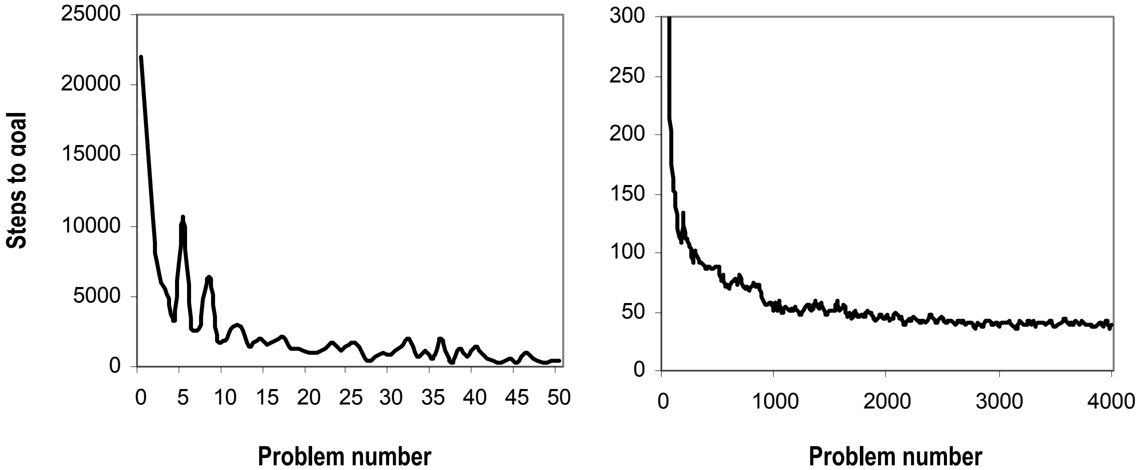


Figure 2. Experimental results in maze1

Each simulation is represented by two graphs. The graph on the left shows the convergence of the number of steps to the goal in the beginning. Here, the learning algorithm shows its ability to roughly sketch out non-optimal but successful behavior. However, after this sketching the optimization of a policy continues very slowly. The second graph on the right shows this slow optimization. The same tendency could be noted on other two graphs.

Table 1 shows summary of all the experiments. It also contains additional information describing the selected maze problems. It should be noted that the number of random steps to the goal from a typical state is extremely large in maze2 and maze3. It is because of irreversible transactions between states in these environments.

The proposed learning algorithm shows very stable ability to form a good policy in each case. However, the experiments have also discovered the drawbacks of the algorithm: slow adaptation and non-optimality of resulting policy. The agent was able to form only a good policy, but not a theoretically optimal one.

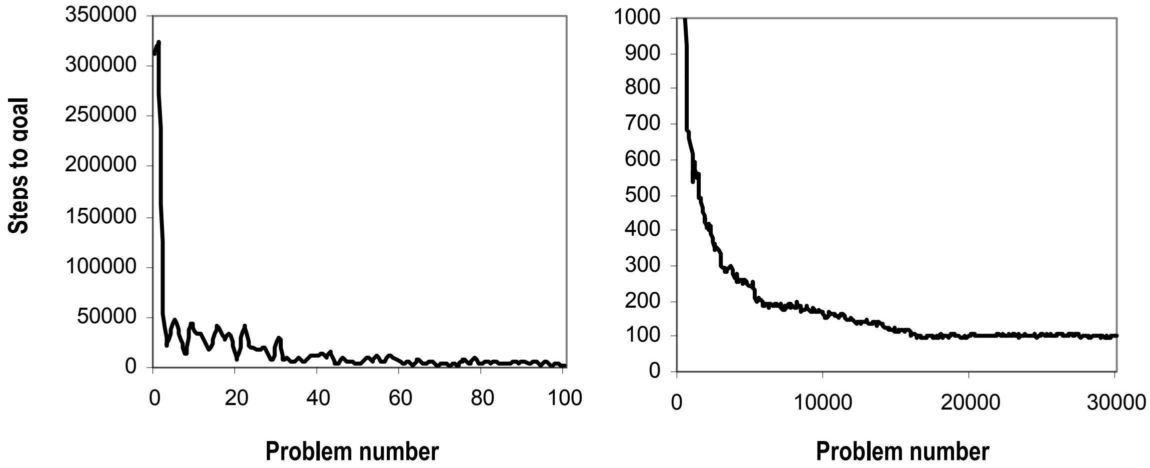


Figure 3. Experimental results in maze2

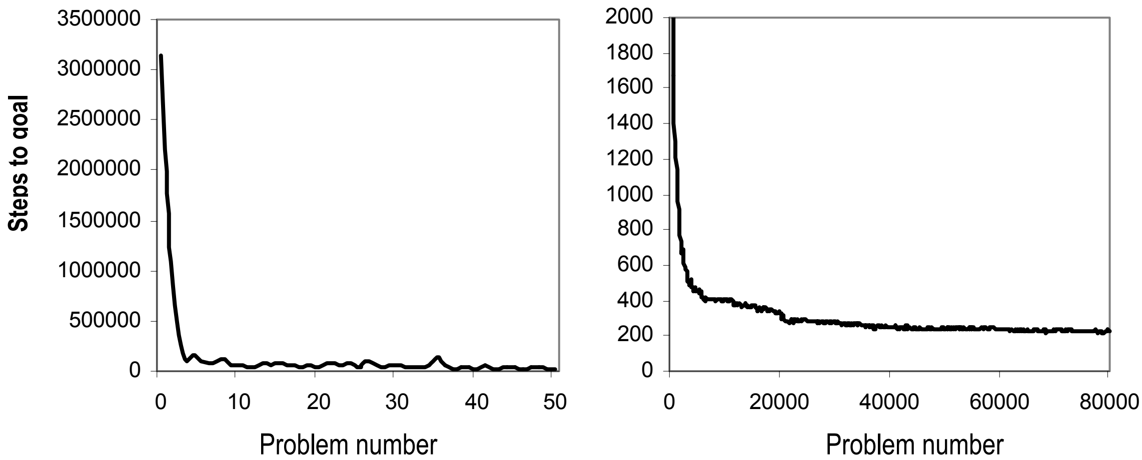


Figure 4. Experimental results in maze3

Table 1. Results summary

		Maze1	Maze2	Maze3
Environment description	The number of perceptions	21	55	89
	The number of positions in the maze	25	56	
	The number of positions in the maze with not unique observations	6	2	32
	The number of states	45	97	208
	The number of belief states	51	98	225
	The number of random steps to the goal from a typical state	$4.1\cdot10^{4}$	$2.3\cdot10^{9}$	$4.3\cdot10^{10}$
	The theoretically optimal number of steps to the goal	12	25	29
Experimental results (on average)	The resulting average steps to the goal	39	97	234
	The time of the first goal achievement	$2.2\cdot10^{4}$	$3\cdot10^{5}$	$3.1\cdot10^{6}$
	The total number of solved problems	$2*10^{3}$	$1.7\cdot10^{4}$	$2\cdot10^{4}$
	The total learning time needed to converge to a good policy	$2.6\cdot10^{5}$	$6.4\cdot10^{6}$	$2\cdot10^{7}$
	During learning the number of steps to the goal was reduced X times	$5.6\cdot10^{2}$	$3\cdot10^{3}$	$1.3\cdot10^{4}$
	The resulting number of steps to the goal is X times bigger than the optimal one	3.25	3.88	8.07

Comparison to other algorithms

Developing a new algorithm, it is important to analyze its place in the context of other algorithms serving for similar purposes. For this aim, it was decided to compare the proposed learning algorithms with other reinforcement learning algorithms. In addition, in order to make such comparison, several criteria were proposed in Table 1. These criteria were used for the comparison presented in Table 2.

It should be noted that the presented evaluation of algorithms is very rough and could not be considered as a fully proved comparison of different algorithms. However, it can help to describe features of the developed algorithm. The advantages of the proposed algorithm are the simple implementation and its stable ability to form a good policy in the extremely complex case when the environment has irreversible transactions and complex temporal dependencies. However, it has also drawbacks: non-optimality and slow adaptation.

Table 2. The description of criteria

Criterion	Description
SI	The implementation is simple.
CT	The computational time taken at each time step is small.
LT	The ability to learn the policy fast.
OPT	The resulting policy is very close to the optimal one.
PO	The ability to overcome the incomplete perception in partly observable environments.
EF	The ability to perform efficient exploration when the environment has many irreversible transactions.
CD	The ability to learn complex temporal dependencies.
PD	The ability to discover perceptual distinctions.
SA	The architecture is not comprehensive or composite.
UL	The number of hierarchy levels is not limited.

Table 3. Comparison to other algorithms

Learning algorithm	SI	CT	LT	OPT	PO	EF	CD	PD	SA	UL
Q-learning [Wat89]	+	+	+/-	+	-	-	-	-	+	-
SSS algorithm [SS2000]	+/-	+	-	+/-	+	-	+/-	-	-	+
HQ-learning [WS98]	+/-	+	-	+/-	+	-	+	-	-	-
USM algorithm [McCallum95]	-	-	+	+	+	-	-	+	+	-
USM + "internal reward for hard-to-reach states" [Pch2003]	-	-	+	+/-	+	+	-	+	+	-
The proposed algorithm	+	+	-	-	+	+	+	-	+	+

Conclusion

In this paper a new hierarchical reinforcement learning algorithm was presented that doesn't exploit any domain-specific knowledge about subtask hierarchy, but automatically builds useful hierarchies. The algorithm was developed with purpose to overcome the combination of three, previously known in reinforcement learning, difficulties: (1) incomplete perception, (2) irreversible transactions and (3) complex temporal dependencies. The key idea of the algorithm is to exploit the observation from the environment not only as context information for action selecting, but also as local, internally selected, goals and subgoals. This makes the agent to be hierarchical reinforcement learner, driven by local goals, that has a native ability of efficient exploration.

The proposed learning algorithm was experimentally investigated in different and very complex maze problems, showing very stable ability to form a good policy in each case. However, the experiments have also discovered the drawbacks of the algorithm: slow adaptation and non-optimality of resulting policy (the agent was able to form only a good policy, but not a theoretically optimal one).

Bibliography

[Bong75] Бонгард, М.М., Лосев, И.С., Смирнов М.С. Проект модели организации поведения – Животное // Моделирование обучения и поведения. – М.: Наука, 1975.

[Chr92] Chrisman, L. Reinforcement Learning with Perceptual Aliasing: The Perceptual Distinctions Approach, In Tenth National Conference on Artificial Intelligence, 1992, pp.183-188.

[Die2000] Dietterich, T.G. Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition, Journal of Artificial Intelligence Research, 13, pp.227-303, 2000.

[KLM96] Kaelbling, L.P., Littman, L.M., Moore, A.W. Reinforcement learning: a survey, Journal of Artificial Intelligence Research, Vol. 4, 1996, pp.237-285.

[Hum96] Humphrys, M. W-learning: a simple RL-based society of mind. Technical report 362, University of Cambridge, Computer Laboratory, 1996.

[Maes95] Maes P., Modeling Adaptive Autonomous Agents, Artificial Life: An overview, MIT Press, 1995.

[McCallum95] McCallum, R.A., Reinforcement learning with selective perception and hidden state (Ph.D. dissertation). Department of Computer Science, University of Rochester, Rochester, NY, 1995.

[McCallum97] McCallum, R.A., "Efficient Exploration in Reinforcement Learning with Hidden State", AAAI Fall Symposium on "Model-directed Autonomous Systems", 1997.

[Mit99] Mitchel, T.H., Machine learning, The McGraw-Hill Companies. Inc., 1999.

[Pch2003] Pchelkin A. Efficient exploration in reinforcement learning based on Utile Suffix Memory, journal “Informatica”, Lithuanian Academy of Sciences, Vol.14., 2003.

[PR97] Parr, R., Russul, S. Reinforcement learning with hierarchies of machines. Advances in Neural Information Processing Systems 9. MIT Press, Cambridge, MA.

[Sut98] Sutton, R.S., Barto, A.G. Reinforcement learning: An introduction. Cambridge, MA: MIT Press, 1998.

[SS2000] Sun, R., Sessions, C. Self-segmentation of sequences: automatic formation of hierarchies of sequential behaviors. IEEE Transactions on Systems, Man, and Cybernetics: Part B Cybernetics, 30(3), 2000.

[Thr92a] Thrun, S.B., Efficient exploration in reinforcement learning, Technical Report CMU-CS-92-102, Carnegie Mellon University, Pittsburgh, PA 15213, 1992.

[Thr92b] Thrun, S.B., The role of exploration in learning control. In DA White & DA Sofge, editors, Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches. New York, NY: Van Nostrand Reinhold, 1992.

[Wat89] Watkins, C. Learning with Delayed Rewards. Ph.D Thesis, Cambrige University, Cambrige, UK.

[Whit91] Whitehead, S.D. Complexity and cooperation in Q-learning, In Proceedings of the Eighth International Workshop on Machine Learning, 1991, pp.363-367.

[Wil85] Wilson S.W. Knowledge growth in an artificial animal, Proceeding of the First International Conference on Genetic Algoritms and Their Applications. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1985, pp.16-23.

[WS98] Wiering, M., Schmidhuber, J. HQ-learning, Adaptive Behavior, vol. 6.2, pp. 219-246, 1998.

Author information

Arthur Pchelkin – PhD. Student**;** Institute of Information Technology, Technical University of Riga; 1, Kalku St., Riga LV-1658, Latvia; e-mail: arturp@inbox.lv

SITUATION CENTERS IN MODERN STATE

M. Slipchenko, A. Kuzemin, M. Sorochan, A. Torojev

Introduction

Development of information technologies and hardware-software means make it possible to integrate various engineering solutions of technical, technological and information nature within the framework of the single system of implementation.

The situation center (SC) represents the complex of hardware-software means for the personal and team work of the managers group. Its main task consists in supporting decision-making on technical and strategic management solutions based on visualization and analytical procession of information.

The SC ensures the support of preparation and decision making in particular:

- elaboration of the versions and recommendations taking into account various conditions and limitations;
- choice or creation of the situation analysis model;
- structurization of the problems and definition of the most informative parameters;
- monitoring of socio-economical and socio-political information.

Being based on the experience of creation and operation of the situation centers in the organs of the state management it is now possible to formulate the main types of solution concerning the contents of hard-ware-software complex of the situation center and operation regimes.

Regimes of holding actions in the SC can be realized in the form of the limited participation of experts or managers and in the form of television conferences (broadened content of participants), they assume joining with the use of telecommunication means of distant representatives from different organizations, enterprises or situation centers.

Thus, information and intelligent resources needed for consideration and elaboration of versions of solutions in a problem situation are consolidated in the situation center. Efficiency of the SC consists in the possibility to include the figurative associative thinking into the active decision-making work.

Designing and operation of the system of the interconnected federal and regional SC create premises for the transition to the fundamentally another technology of operation of the state management of all levels, namely, to the usage of the new “generation” of the decision-making methods directed to the work of experts and persons making decisions in fuzzy multivariant conditions.

Application of information technologies and systems in the practice of the situation centers functioning

Information technologies and systems make the basis of the SC efficient functioning. In addition to that their practical application is based on the definition of the objects’ domains which in the state management are: the national security, economy, financial activity, social sphere, external and internal policy, branches of industry, fuel – power complex, transport, agriculture, development of regions, objective programs etc.

Means of information collection and storage is realized in the forms of based and data banks, knowledge bases as well as the systems of telecommunication information action.

Forms of information storage represent segments with databases of the problem and objective orientation as well as booths of data with the selective data in terms of the users orientation. The segments’ database and booths are universally connected with the operation and special databases in the federal and regional structures. They are united in the global network of data collection arriving from different sources.

In the context of development of network technologies it became urgent to ensure the users interface with the distant (external) sources of information. One of the efficient ways to solve the emerging problems is creation of the unique mechanism of access to the internal and external resources in the form of the corporative information portal.

Information portal integrates such internal applications as: e-mail, access to databases, OLAP, Data Mining and DSS, systems of documents management with the external application (Internet news services and users' Web-nodes). Modern intelligent information technologies used in the SC can be represented by four groups of technologies: operative analysis, intelligent analysis, simulation and decision-making solutions support.

Technology of operative analysis of data makes it possible to estimate the state of the processes being observed to reveal and range the causes of significant variations to forecast development of the processes.

Recently a wide-spread occurrence was obtained by analytical systems based on OLAP technologies (OLAP – on-line analytical processing – operative analytical data processing), which make it possible to group the data into representation of information in the form of N-dimensional cube. Such a technology gives the possibility to an analyst to receive "multidimensional reports" from miscellaneous sources of information and form the necessary sampling in the form of different data cut.

Intelligent analysis of data ensures automated search of earlier unknown regularity in the databases of the information fund. The use of the acquired knowledge allows so simplify significantly the procedure of the informative analysis for the analysts and to increase efficiency of the models' design by experts. Methods of the artificial intelligence make the basis of the intelligent analysis.

Simulation of the decision-making process makes it possible to give a quantitative estimation and perform quantitative analysis of results of the made decisions.

In these technologies oriented to the users from the organs of the state power the models are used classified as the models of socio-economical processes, models of socio-political processes as well as the models of extraordinary situations.

Thus, in the socio-economical sphere the use of the models of regression analysis makes it possible to predict dynamics of macro economical indices, development of different branches of industry and agriculture, to perform comparative analysis of socio-economical situation of regions in Russian Federation, to estimate their investment attractiveness etc.

In the socio-political sphere, for example, the model calculation is widely used for procession and analysis of data of the sociological enquiries.

In the process of preparation of decisions in the extraordinary situations simulation models are widely used, they allow to analyze the development of the extraordinary situations, to estimate their consequences and calculate the required resources needed for elimination of resources damage.

Decision-making support systems are singled out separately; they are functionally oriented to the preparation of the analytical reports and documents, performance of the group expertises, development of recommendations and decision versions ranging. In these systems they used mainly methods of expert estimates, models of group estimates, method of analysis of hierarchy (method of pair comparisons), method of alternative decisions synthesizing.

Information-analytical systems input into software-hardware complex of the SC significantly extend its functional possibilities. In this framework of the highest organs of the state power the most required are the general mathematical information-analytical systems (IAS).

The means of the information presentation unite the following types of technologies:

- Cartography of the problem situations and objects of decisions;
- Structuring of fuzzy ideas and decision hypothesis;
- Multimedia imagery of the situation dynamics.

Problems of information security

Provision of the information security is one of the most important problems in application of the information technologies in the practice of the state management. Organization of the systems' functioning in the modern communication means in conjunction with the necessity to perform the requirements on the user service significantly complicates this problem solution.

Two types of information struggle should be singled out, namely, information-psychological and information-technical ones. The main objects of impact and security in information technical struggle are information-engineering systems (communication systems, telecommunications system, radio electronic means etc).

Realization of information security hazard can result in a serious and, in a number of cases, catastrophic consequences; the main of them are as follows:

- violation of the state establishments, social organizations and institutions activity;
- variation of the individual or mass consciousness, moral and political, social and psychological chi mate in the society;
- infringement of the state national interests due to the drain of the most important information.

Prospects on the information technologies development and proposals on the situation centers development

Today the level of development of the information technologies makes it possible to envisage confidently the possibilities of creating the system of “electronic power” in the near future. The essence of the efficient control and the processes of decision-making management at the first stage and further in formation of the complex system of control of socio-economical and socio-political processes in the context of the state management.

Under conditions of the global information of the society the role and functions of management transform, this, respectively, is reflected on the realization of economic and social policy as well as on the support of democratic institutes of power. At present it becomes evident that in future the efficiency of the process of management will depend even greatly on the quality of information. In this case the main methodical directions of rising quality of the information support of the state management in the context of the considered problem are, at our point of view, as follows:

- development of methodological principles for introducing the situation centers as means of the system integration of the intelligent information technologies into practice of the state management of the federal and regions levels;
- development of the methodology for consolidation of the information resources and clear definition of their propagation limits, creation and development of the information funds on the basis of the distribution calculations methodology and data storage;
- creation of the decision-making system on the basis of the modern intelligent information means of data procession;
- development of a wide class of the socio-economical and socio-political processes models;
- orientation in development and application to the management of the intelligent information technologies aimed to creation of the systems of “electronic power”;
- creation of the efficient technologies of the information security oriented to application to information-analytical systems of the state management.

Conclusion

Results of the imitative simulation systems’ use in the situation centers are given in this work. Majority of the program complexes, used in the world for the economical, political, and financial simulation, are based on the methods of the so called system dynamics. The latter one, in its turn, uses the fuzzy cognitive maps’ apparatus offered by Kosko at the beginning of the eighties and used for the first time in the field conditions during the political crisis in South Africa.

Authors' Information

Slipchenko M.I.: Prof., Prorector of Kharkov National University of Radio Electronics

Kuzemin A.Ya.: Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, (Ukraine), kuzy@kture.kharkov.ua

Sorochan M.V.: post-graduate student of Kharkov National University of Radio Electronics, msorc@ua.fm

Torojev A.A.: General Director of the Joint Russian-Kirghiz Venture “Computing Technique and Automation Means”

AN INTERACTIVE METHOD OF LINEAR MIXED INTEGER MULTICRITERIA OPTIMIZATION

V. Vassilev, K. Genova, M. Vassileva, S. Narula

***Abstract**: The paper describes a learning-oriented interactive method for solving linear mixed integer problems of multicriteria optimization. The method increases the possibilities of the decision-maker (DM) to describe his/her local preferences and at the same time it overcomes some computational difficulties, especially in problems of large dimension. The method is realized in an experimental decision support system for finding the solution of linear mixed integer multicriteria optimization problems.*

***Keywords**: linear mixed integer multicriteria optimization, interactive methods, decision support systems.*

1. Introduction

The reference point interactive methods and the classification-based interactive methods are the most widely spread methods [Gardineer and Vanderpooten, 1997], [Miettinen, 1999], [Vassilev et al., 2003] for solving linear problems of multicriteria optimization. This kind of interactive methods is also used [Climaco et al., 1997], [Vassileva, 2001] for solving linear mixed integer problems of multicriteria optimization.

Single-criterion problems of linear programming are used in the interactive methods solving linear problems of multicriteria optimization. They belong to the class of P-problems [Garey and Johnson, 1979] and therefore the time for solving a single-criterion problem does not play a significant role in the interactive methods of linear multicriteria optimization. In the development of the interactive methods, main attention is paid to the possibilities offered to the DM to describe his/her preferences. Single-criterion problems of mixed integer programming are applied in the interactive methods solving linear mixed integer problems of multicriteria optimization. They are NP-hard problems and hence the time for single-criterion mixed integer problems solution has to be obligatory taken into account when designing interactive methods.

A learning-oriented interactive method is proposed in the present paper on the basis of new linear mixed integer classification-based scalarizing problems, intended to solve linear mixed integer problems of multicriteria optimization. The method increases DM's possibilities to describe his/her local preferences and also overcomes some computational difficulties, connected with single-criterion mixed integer problems solving.

2. Problem formulation

The linear mixed integer problem of multicriteria optimization (denoted by **I**), can be formulated as follows:

(1) $\text{"max"}\{f_k(x)\}, \qquad k \in K$

subject to:

(2) $\sum_{j\in N} a_{ij}x_j \le b_i, \qquad i \in M$

(3) $0 \le x_j \le d_j, \qquad j \in N$

(4) x_j – integers, $j \in N'$; $N' \subseteq N$,

where $f_k(x), k \in K$ are linear criteria (objective functions), $f_k(x) = \sum_{j=N} c_j^k x_j$, "max" denotes that all the objective functions should be simultaneously maximized; $x = (x_1, x_2 ..., x_j, ...x_n)^T$ is the variables vector; $K = \{1,2,..., p\}$, $M = \{1,2,...,m\}$, $N = \{1,2,...,n\}$ and $N' \subseteq N$ are sets of the indices of the linear criteria (objective functions) of the linear constraints, of the variables and of the integer variables respectively. Constraints (2)-(4) define the feasible region X_1 for the variables of the mixed integer problem.

Several definitions will be introduced for greater precision.

Definition 1: The solution x is called an efficient solution of problem (**I**), if there does not exist any other solution $\bar{x}$, so that the following inequalities are satisfied:

$f_k(\bar{x}) \geq f_k(x),$ for every $k \in K$ and

$f_k(\bar{x}) > f_k(x),$ for at least one index $k \in K$.

Definition 2: The solution x is called a weak efficient solution of problem (**I**), if there does not exist another solution $\bar{x}$ such that the following inequalities hold:

$f_k(\bar{x}) > f_k(x),$ for every $k \in K$.

Definition 3: The solution x is called a (weak) efficient solution, if x is either an efficient or a weak efficient solution.

Definition 4: The vector $f(x) = (f_1(x),..., f_p(x))^T$ is called a (weak) non-dominated solution in the criteria space, if x is a (weak) efficient solution in the variables space.

Definition 5: A near (weak) non-dominated solution is a feasible solution in the criteria space located comparatively close to the (weak) non-dominated solutions.

Definition 6: A current preferred solution is a (weak) non-dominated solution or near (weak) non-dominated solution, if selected by the DM at the current iteration.

Definition 7: The most preferred solution is the current preferred solution, which satisfies the DM to the highest extent.

3. Scalarizing problems

The linear mixed integer problems of multicriteria optimization do not possess a mathematically well-defined optimal solution. That is why it is necessary to choose one of the (weak) non-dominated solutions, which is most appropriate with reference to DM's global preferences. This choice is subjective and it depends entirely on the DM.

The interactive methods are the most widely used methods for solving linear mixed integer problems of multicriteria optimization [Climaco et al., 1997], [Vassileva, 2001]. Every iteration of such an interactive method consists of two phases: a computation and a decision one. One or more non-dominated solutions are generated with the help of a scalarizing problem at the computation phase. At the decision phase these non-dominated solutions are presented for evaluation to the DM. In case the DM does not approve any of these solutions as a final solution (the most preferred solution), he/she supplies information concerning his/her local preferences with the purpose to improve these solutions. This information is used to formulate a new scalarizing problem, which is solved at the next iteration.

The type of the scalarizing problem used lies in the basis of each interactive method. The scalarizing problem is a problem of single-criterion optimization and its optimal solution is a (weak) non-dominated solution of the multicriteria optimization problem. The classification-based scalarizing problems are particularly appropriate in solving linear mixed integer multicriteria optimization problems, because they enable the decrease of the computational difficulties connected with their solution as well as the increase of DM's possibilities in describing his/her local preferences and also lead to reduction of the requirements towards the DM in the comparison and evaluation of the new solutions obtained. In the scalarizing problems suggested in this chapter, the DM can present his/her local preferences in terms of desired or acceptable levels, directions and intervals of alteration in the values of separate criteria. Depending on these preferences, the criteria set can be divided into seven or less criteria classes. The criterion $f_k(x), \quad k \in K$ may belong to one of these classes as follows:

$k \in K^{>}$, if the DM wishes the criterion $f_k(x)$ to be improved;

$k \in K^{\geq}$, if the DM wishes the criterion $f_k(x)$ to be improved by any desired (aspiration) value Δ_k;

$k \in K^{<}$, if the DM agrees the criterion $f_k(x)$ to be worsened;

$k \in K^{\leq}$, in case the DM agrees the value of the criterion $f_k(x)$ to be deteriorated by no more than δ_k;

$k \in K^{><}$, if the DM wishes the value of the criterion $f_k(x)$ to be within definite limits with respect to the current value f_k, $(f_k - t_k^- < f_k(x) \leq f_k + t_k^+)$;

$k \in K^{=}$, if the current value of the criterion $f_k(x)$ is acceptable for the DM;

$k \in K^{0}$, if at the moment the DM is not interested in the alteration of the criterion $f_k(x)$ and this criterion can be freely altered.

In order to obtain a solution, better than the current (weak) non-dominated solution of the linear mixed integer problem of multicriteria optimization, the following Chebyshev's scalarizing problems can be applied on the basis of the implicit criteria classification, done by the DM. The first mixed integer scalarizing problem [Vassileva, 2000] called *DAL* (desired and acceptable level) has the following type:

To minimize:

(5) $$S(x) = \max\left[\max_{k \in K^{\geq}}\left(\bar{f}_k - f_k(x)\right)/\left|f_k^{'}\right|, \max_{k \in K^{\leq}}\left(f_k - f_k(x)\right)/\left|f_k^{'}\right|\right]$$

under the constraints:

(6) $$f_k(x) \geq f_{k,}\ k \in K^{=},$$

(7) $$f_k(x) \geq f_k - \delta_k,\ k \in K^{\leq},$$

(8) $$x \in X_1,$$

where:

f_k, $k \in K$ is the value of the criterion $f_k(x)$ in the current preferred solution;

$\bar{f}_k = f_k + \Delta_k$, $k \in K^{\geq}$ is the desired level of the criterion $f_k(x)$;

$f_\kappa^{'}$, $k \in K$ is a scaling coefficient:

$$f_k^{'} = \begin{cases} \varepsilon, \text{ if} \left|f_k^{'}\right| \leq \varepsilon \\ f_{k,} \text{ if} \left|f_k^{'}\right| > \varepsilon \end{cases},$$

where ε is a small positive number.

DAL scalarizing problem has three properties, which allow to a great extent the overcoming of the computational difficulties, connected with its solving as a problem of integer programming and also decrease DM's efforts in the comparison of new solutions. The first property is connected with this, that the current preferred integer solution of the multicriteria problem (found at the previous iteration), is a feasible integer solution of *DAL* problem. This facilitates the exact as well as the approximate algorithms for solving *DAL* problem, because they start with a known initial feasible integer solution. The second property is connected with the fact, that the feasible region of *DAL* problem is a part of the feasible region of the multicriteria problem (**I**). Depending on the values of the parameters $\Delta_k\ / k \in K^{\geq}$, $\delta_k\ / k \in K^{\leq}$ the feasible region of *DAL* problem can be comparatively narrow and the feasible solutions in the criteria space, found with the help of approximate integer programming algorithms, may be located very close to the non-dominated surface of the multicriteria problem (**I**). The third property comprises DM's possibility to realize searching strategy of "not big profits – small losses" type. This is due to the fact, that such optimal solution is searched for with the help of *DAL* problem, which minimizes Chebyshev's distance between the feasible criteria set and the current reference point, the components of which are equal to the wished by the DM values of the criteria being improved and to the current values of the criteria being deteriorated. The (weak) non-dominated solution obtained and the current solution are comparatively close and the DM can easily make his/her choice.

The classification-oriented scalarizing problems are appropriate in solving integer problems of multicriteria optimization, because the computational difficulties, connected with their solving are decreased with their help and the requirements towards the DM in the comparison and evaluation of the new solutions obtained, are diminished. From a viewpoint of the information, required by the DM in new solutions seeking in this scalarizing problem, the DM has to define the desired or acceptable levels for a part or for all the criteria. With the help of the scalarizing problem below described, called *DALDI* (desired or acceptable level, direction and interval), the DM is able to present his/her local preferences not only by desired and acceptable levels, but also by desired and acceptable directions and intervals of alteration in the values of separate criteria. The mixed integer scalarizing problem *DALDI* has the following type:

Minimize:

(9) $$S(x) = max\left[\max_{k\in K^{\geq}}\left(\bar{f}_k - f_k(x)\right)/\left|f_k'\right|, \max_{k\in K^{<}\cup K^{\leq}}\left(f_k - f_k(x)\right)/\left|f_k'\right|\right] + \max_{k\in K^{>}}\left(f_k - f_k(x)\right)/\left|f_k'\right|,$$

under the constraints:

(10) $$f_k(x) \geq f_{k,}\ k \in K^{>} \cup K^{=},$$

(11) $$f_k(x) \geq f_k - \delta_k,\ k \in K^{\leq},$$

(12) $$f_k(x) \geq f_k - t_k^-,\ k \in K^{><},$$

(13) $$f_k(x) \leq f_k + t_k^+,\ k \in K^{><},$$

(14) $$x \in X_1,$$

The scalarizing problem *DALDI* has characteristics similar to *DAL* scalarizing problem, but still there are two differences between them.

The first difference consists in this, that *DALDI* scalarizing problem gives greater freedom to the DM when expressing his/her local preferences in the search for a better (weak) non-dominated solution. Besides desired or acceptable values of a part or of all the criteria, the DM has the possibility to set also desired or acceptable directions and intervals of change in the values of some criteria The second difference between *DAL* and *DALDI* scalarizing problems concerns the possibility to alter their feasible sets (make them "narrower"), so that the feasible solutions are positioned close to the non-dominated (efficient) solutions of the multicriteria problem The more the criteria are, which the DM wishes to be freely improved or freely deteriorated, the smaller this possibility is. The narrow feasible regions of the scalarizing problems *DAL* and *DALDI* enable the successful application of approximate single-criterion algorithms, which is especially important when these problems are integer. It should be noted that scalarizing problem *DAL* is better than scalarizing problem *DALDI* in this aspect.

DAL and *DALDI* problems are nonlinear mixed integer programming problems [Wolsey, 1998]. Equivalent linear mixed integer programming problems can be constructed, [Vassileva, 2000], [Vassilev et al., 2003] with the help of additional variables and constraints

4. GAMMA-I1 interactive method

On the basis of scalarizing problems *DAL* and *DALDI*, a classification-oriented interactive method, called **GAMMA-II** is proposed for solving linear mixed integer programming problems of multicriteria optimization. The problems of mixed integer programming are NP-problems, i.e. the time for their exact solution is an exponential function of their dimension. That is why, in solving integer problems, particularly problems of larger dimension (above 100 variables and constraints), some approximate methods are used [Vassilev and Genova, 1991], [Pirlot, 1996]. Since finding a feasible solution is as difficult as finding an optimal solution, the approximate integer methods in the general case do not guarantee the finding of an optimal integer solution and of an initial feasible integer solution too. If the initial feasible integer solution is known and the feasible region is comparatively "narrow", then with the help of the approximate integer methods some satisfactory and in part of the cases optimal integer solutions could be found. The scalarizing problems *DAL* and *DALDI* have known feasible initial integer solutions.

DALDI scalarizing problem allows enlargement of the information with the help of which the DM can set his/her local preferences. This information expansion leads to the extension of the feasible set of criteria alteration in the criteria space and of the integer variables in the variables space. Hence, the approximate integer solutions of *DAL* problem (obtained with the help of an approximate integer method) are located closer to the non-dominated (efficient) set of the multicriteria problem, than the approximate solutions of *DALDI* problem. Therefore, if solving linear mixed integer problems of (*I*) type of large dimension, when the scalarizing problems have to be solved approximately in order to reduce the waiting time for new solutions evaluated by the DM, it is better to use *DAL* scalarizing problem than scalarizing problem *DALDI*.

Two different strategies are applied in the development of **GAMMA-I1** interactive method in the process of searching for new solutions that are evaluated. The first strategy, called integer strategy, consists in seeking a (weak) non-dominated integer solution at each iteration by exact solution of the corresponding linear mixed integer scalarizing problem. The second strategy, called approximate integer strategy, comprises searching for near (weak) non-dominated integer solutions at some iteration, approximately solving a respective linear mixed integer scalarizing problem. During the learning phase and in problems of large dimension up to the very end, only near (weak) non-dominated solutions can be looked for.

The interactive **GAMMA-I1** method is designed to solve linear mixed integer problems of multicriteria optimization. The two strategies above described are realized in the method during the search for new solutions for evaluation in order to overcome the computational difficulties (particularly in solving problems of large dimension). The method is oriented towards learning and the DM has to determine when the most preferred solution is found. The algorithmic scheme of **GAMMA-I1** interactive method includes the following basic steps:

Step 1. An initial near (weak) non-dominated solution is found, setting $f_k = 1$, $k \in K$ and $\bar{f}_k = 2$, $k \in K$ and solving *DAL* problem.

Step 2. The current (weak) non-dominated solution or near (weak) non-dominated solutions obtained are presented for evaluation to the DM. If the DM evaluates and chooses a solution that satisfies his/her global preferences, Step 6 is executed, otherwise – Step 3.

Step 3. A question is set to the DM what new integer solution he wishes to see – a (weak) non-dominated or a near (weak) non-dominated solution. Step 5 is executed in the first case and Step 4 - in the second.

Step 4. The DM is asked to define the desired or feasible levels of the values of a part or of all the criteria. Scalarizing problem of *DAL* type is solved and then goes to Step 2.

Step 5. The DM is requested to define the desired or feasible levels, directions and intervals of alteration in the values of a part or of all the criteria. Scalarizing problem of *DALDI* type is solved and then goes to Step 2.

Step 6. Stop the process of the linear mixed integer multicriteria problem solving.

In **GAMMA-I1** interactive method the DM controls the dialogue, the computing process and the conditions for cancelling the process of linear mixed integer multicriteria problem solution.

5. Conclusion

The interactive **GAMMA-I1** method is realized in the experimental software system **MOLIP**, developed at the Institute of Information Technologies of the Bulgarian Academy of Sciences. This system is designed for interactive solution of linear and linear mixed integer multicriteria optimization problems with different number and type of the criteria, with different number and type of the variables and constraints. **MOLIP** system functions in the environment of Windows 98 and higher versions and may serve for learning purposes, as well as for the solution of different applied problems. Our experimental results confirm that the computational effort and time are reduced considerably using heuristic integer algorithms in the learning phase and when solving large problems.

Bibliography

[Climaco, et al., 1997] J. Climaco, G. Ferreira, and M. E. Captivo. Multicriteria Integer Programming: An Overview of the Different Algorithmic Approaches In: Multicriteria Analysis (J. Climaco, Ed.). Springer, 248-258, 1997.

[Gardiner and Vanderpooten, 1997] L.R.Gardiner and D.Vanderpooten. Interactive Multiple Criteria Procedures: Some Reflections. In: Multicriteria Analysis (J. Climaco, Ed.). Springer, 290-301, 1997.

[Garey and Johnson, 1979] M. Garey and D. Johnson. Computers and Intractability: A Guide to the Theory of NP-completeness. W. H. Freeman, San Francisco, 1979.

[Miettinen, 1999] K. Miettinen. Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston, 1999.

[Pirlot, 1996] M. Pirlot. General Local Search Methods. European Journal of Operational Research, 92, 3, 493-511, 1996.

[Vassilev et al., 2003] V. Vassilev, K. Genova, M. Vassileva and S. Narula. Classification-Based Method of Linear Multicriteria Optimization. International Journal on Information Theories and Aplications, vol.10, 3, 266-270, 2003.

[Vassilev and Genova, 1991] V. Vassilev and K. Genova. An Algorithm of Internal Feasible Directions for Linear Integer Programming. European Journal of Operational Research, 52, 203-214, 1991.

[Vassileva, 2000] M. Vassileva. Scalarizing Problems of Multiobjective Linear Programming Problems, Problems of Engineering Cybernetics and Robotics, 50, 54-64, 2000.

[Vassileva, 2001] M. Vassileva. An Interactive Algorithm of the Multicriteria Linear Integer Programming. Problems of Engineering Cybernetics and Robotics, 52, 17-25, 2001.

[Wolsey, 1998] L. Wolsey. Integer Programming. Wiley – Interscience, 1998.

Author information

Vassil Vassilev - Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: vvassilev@iinf.bas.bg

Krasimira Genova - Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: kgenova@iinf.bas.bg

Mariyana Vassileva - Institute of Information Technologies, BAS, Acad. G. Bonchev St., bl. 29A, Sofia 1113, Bulgaria; e-mail: mari@iinf.bas.bg

Subhash Narula - School of Business, Virginia Commonwealth University, Richmond, VA 23284- 4000, USA; e-mail: snarula@vcu.edu

DIGITAL CREATIVITY: ADVANTAGES, PROBLEMS, RESPONSIBILITIES

V. Velev

***Abstract**: The paper focuses on the rapid development of the digital culture and the challenges it imposes to human creativity. It analyses e-learning, digital entertainment, digital art and the issues of creativity and improvisation. It also presents a classification of the levels in the creative structure including hardware and software tools; product developers; creators and end users. Special attention is paid to the advantages of the new digital culture and the responsibilities of all people who create it or use it. We conclude that more attention should be paid to the threats and to ways of boosting positive creativity in the various fields of application of information and communication technologies.*

***Keywords**: digital art, creativity, improvisation*

Introduction

The rapid development of information and communication technologies (ICT) implied important changes in the human lives. The earliest uses of computers were very narrow-profiled and applied in specific sectors and professions. Nowadays, ICT accompany people from early childhood in the educational system, at work, in the societal organisation, in everyday life activities, and at leisure. This trend seems so natural that one important shift was left somewhat behind the focus: the function of computers changed. Starting with support for humans in routine tasks, nowadays computers became in the centre of wide range of tasks, which involve human creativity. This is obvious in areas related to computer art, but is equally important in entertainment and educational applications, which form the fabrics of the human personality, including the creativity.

The basic aim of this paper is to increase the awareness of creativity related issues in the modern computer world and its influence on the human personality.

We provide an analysis of the typical areas influenced by the new digital culture: e-learning, digital entertainment, digital art and discuss the issues of creativity and improvisation in the digital environment.

We continue with a presentation of our view on the digital culture ingredients and their inner dependencies. The basic ingredients of the digital culture include hardware and software tools, developers and users. The users basically form two groups – consumers and creators. One would expect that one consequence from the use of the ICT would be increased creativity and improvisation. This matter is not studied yet and we hope that this paper will raise the interest to it.

Finally, we present a SWOT (Strengths-Weaknesses Opportunities-Threats) analysis of the digital culture. Our basic conclusion is that more attention should be paid to raise the awareness in the threats and to boost positive creativity in order to support the development of the unique distinctive creative nature and potential of the human personality and to save the human values in the society as a whole through the various fields of application of ICT.

Forms of Life of Digital Culture

Nicholas Negroponte, the founder and the director of the Media Laboratory at the Massachusetts Institute of Technology, wrote in the introduction to his book *Being Digital*: "Computing is not about computers anymore. It is about living" (p. 6, [Negroponte 1996]).

The digital technologies led to the creation of a new world, a 'different reality' that we often call 'virtual space' or 'virtual environment' (VE) The technological developments, which make possible human beings to immerse into it, are becoming more and more mature. Almost everything which one could ever imagine could be modelled inside of the virtual environment: one can 'live' and 'do' there, i.e. one can simulate all types of human activity.

There are several application fields, which are particularly strongly influenced by the idea of virtual environment and human immersion:

- E-learning
- Digital entertainment
- Digital art and its genres

E-learning

E-learning is defined as "The use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services as well as remote exchanges and collaboration."[26]This definition puts the emphasis on two basic issues: specific technologies (multimedia, Internet) and on the process of learning. In this field, there is enormous space for creativity. We are moving to the development of a virtual community, which shares a common educational content.

However, the e-learning should not be overestimated. Even if fields like linguistics or pure information, there is a small share, which can be filled in only by the real practice. Without it nowadays and in the future the process of learning could not be completed adequately. This is not only due to the fact that devices simulating physical sensation (haptic devices, olfactory interfaces, immersive visual interfaces) are still not perfect, but because of the impossibility to simulate and substitute the human presence. The contact with the other person, no matter whether he is a supervisor, partner or patient, and most over the vivid human feedback could not be replaced even in the most advanced digital technologies. To many people this statement could seem old-fashioned, but we should not neglect that in each human sphere there is a certain amount of knowledge, which can be transmitted only via personal contact.

Michael Polanyi, the creator of the theory of the personal knowledge, concludes that there are two types of human knowledge – the first one is obvious and logical, and the other one is hidden, peripheral [Polanyi 1962]. The second one is acquired through studying of the internal structure of the object of study and this type of knowledge can be transmitted only through personal contact between the teacher and student. The obvious knowledge is transferred logically and verbally. The hidden knowledge is transmitted only directly in the contact of teacher and student.

The informal personal knowledge could be transmitted only within the process of personal contact. Thomas Kuhn investigates the phenomenon why good students most usually are not becoming successful researchers [Kuhn 1962]. He suggested that this is due to their maturing basically through reading textbooks and the lack of personal contact and real practice.

Of course, we are not denying the role, place and the meaning of the e-learning. It is a new kind of enriching the forms of education nowadays. Being accessible, financially profitable and safe in the cases of simulating of critical situations, it is unique and can give us a lot, and shorten the periods of training. In the future, it could become obligatory way of knowledge acquisition. But we should understand that it could be necessary, but not sufficient form of training and it is better to use it as a step to the last phase of training: the real practice in our "material" world.

Digital Entertainment

In the USA the number of sales of games already overcomes book sales. In the UK the games are exceeding by 80% video rentals [Ross et al, 2003]. This illustrates the great power of games industry and opens the question how it is actually utilised by the society.

The possibilities for building edutainment (educational entertainment) products involving computer games in the field of cultural heritage are studied in [Sotirova, 2004]. This paper shows clearly that developing games with cultural and historical content is a difficult process and needs special measures from governments or other organizations which would support it – otherwise the companies would normally go for the development of more cost-effective products, which on the long-run do not contribute to the balanced development of the young generation.

In the digital entertainment we find the combination of spectacular scenery and personal involvement, 'real' participation in the virtual life. The attractive power of games is in immersing in worlds, which exist only in the imagination. While in the daydreams one can model the reaction of the opposite side and the outcome of the

[26] http://www.elearningeuropa.info, last visited on 22 April 2004.

events, here one meets other people or artificial intelligence, which are beyond his power. Here the opposite side is independent, acts on its own will, has its own logic and the outcome is not predictable. This builds the adventure spirit and makes the emotional experience especially strong.

Digital art and its genres

The search for the term "Digital art" returned 1,580,000 results on 22 April 2004. There are websites, which provides tens of thousands of pages of galleries and advice for the digital artist. A typical one is 'The Digital Art Community"[27] which offers tutorials for the people willing to become digital artists, e.g. 3D modelling/Animation, anatomy and Figure Drawing, Concept and Design, Digital Illustration and Painting, Photoshop Effects and Tips. On the one hand, we witness enormous amount of resources on this topics and of works called by their authors 'digital art'. On the other hand, the advice, which is given, most often is targeted to acquiring technical skills. Is this what makes the artist? Where one could learn creativity or improvisation?

Now in the digital environment we have plethora of new art directors, new scriptwriters, new artists – this is the new art. However, the aesthetic, moral and spiritual laws which are valid for the traditional art are valid here too; the same codex and the same principles because the essence of the art is the same, only the tools are different. But we should take into account that the new art compared to the traditional one is different in the sense of power of effect on the human consciousness. The power of art, the large media scale, the magic of the drugs, all of these are collected here, but strengthened several times. Its power of effect is tenfold bigger than that of the cinema and visual arts; hundredfold than that of the theater and thousandfold more than that of the literature, because it collects all of them together.

Digital creativity and improvisation

A substantial part of the software tools – these for graphical design, manipulation of 3D objects, space and animation, as well as the specialised applications for creative modelling, e.g. in fashion, agent and avatar technologies, stimulate the development of the creative nature of a growing number of people. Everyone could find something, which suits his personal taste; his individuality and interests.

For example, People Putty is a program from Haptek[28] that allows one to make his own interactive 3D characters changing skins, shapes, emotions and accessories. The game developers are also creators. Every digital maniac could create a game and settle it with creatures without being trained to draw, to write or to compose music. People can make different things: their own greeting cards, electronic journals, web pages and thus have numerous opportunities to create.

The most incredible is that everyone can try whatever he likes – to draw, to design, to fight, to test crazy ideas, to make impossible dreams become true. The most substantial advantage is that everyone can **improvise**, i.e. to entertain himself, to reveal **creativity**. Everyone could be an artist, architect, musician, designer, builder, warrior, politician, king, and even God! This is so easy, safe and without any responsibility! And that is the most dangerous side of it. False creative self-confidence is being formed and the need for preceding professional training disappears. This leads to invasion of unprofessionalism in the art. Creation of clichés suffocates originality and creative thinking. Every work despite its artistic value becomes public and natural censorship regarding artwork and artists is lacking. All this leads to decay of the esthetical criteria and expectations of the user.

The improvisation in the digital environment is basically taking the form of a game and creative entertainment, while the professional creative improvisation has a preliminary defined topic, clear idea and a structural skeleton around which the improvisation is built. The topic and the idea are leading in the choice of means of expression and the connections within the skeleton. The creative improvisation follows one general idea and a message. The difference in the digital improvisation is in the level of consciousness and preliminary planning. Digital improvisation can cover the range of irresponsible entertainment to the conscious creative act, but this depends on the user. Thus, the users can range from users consumers to users creators.

---

[27] http://www.gfxartist.com, last visited on 22 April 2004.

[28] http://www.haptek.com/, last visited on 16 May 2004.

The Digital Culture Ingredients

After presenting the fields, which are most strongly influenced by the modern digital culture, we would like to discuss what are its current ingredients and bearers. We start with the traditional grouping to *hardware* and *software* and add two core groups of people connected with the digital culture: *developers* and *users.*

Hardware

The computer hardware is the presupposition for the functioning of a specific application. It could facilitate or make more difficult the execution of a specific task. The current struggle of companies in this field is for building devices, which are *easy to use, do not require special education and effort*, and are *portable* (see the section on Human-computer Interaction in [Ross et al., 2003]). Additional trend in the recent years is that the specialisation of devices blurs and we use more and more *multifunctional devices*, e.g. mobile phones, which include game stations, digital cameras, audio recorders and players.

One trend of special importance in the current hardware developments is the increased role of *virtual environments (VE)* opposed to *augmented reality.* One important concept related to VEs is the human *immersion* (see, e.g., the section on Virtual reality in [Ross et al., 2003])

Virtual Environments

The basic role of VE technology is to provide tools for the development and use of computer-generated artificial environments. It is believed that thus users will interact with the environment in natural and easy ways. In this respect the interface issues play crucial role. Probably the basic reason for interest of professionals in VE systems is that they contribute to work on new application areas, which were too expensive or dangerous before (like underwater and space work, hazardous environments, simulation of natural calamities). In addition it is considered that well-designed VEs would provide more intuitive metaphors for human-computer interaction. VEs quickly become one of the basic ingredients of computer games.

VEs can be divided into three groups differing on the sense of immersion they provide: *non-immersive, semi immersive* and *fully immersive* [Ross et al, 2004]. Immersion is measured by the power of attention, which the user focuses on the activity and his/her sense of presence in the environment. The sense of immersion is directly connected with the number of senses involved in the user work. This goes beyond the visual and audio communication channels, which is traditionally used in human-computer interaction and involves devices, which use the sense of touch, sense of smell and movement.

For the visual channel, despite the display characteristics, other factors contributing to the sense of immersion include image quality (including number of dimensions) and the speed and level of interactivity of the system.

Non-immersive systems are the weakest ones. The VEs are simulated on a desktop computer with a standard high-resolution display. Keyboard, mouse, joysticks are employed for interaction with the system. Data gloves often are used in such systems to involve the sense of touch in the user work.

Semi-immersive systems provide better sense of immersion. This is achieved through the use of large screens of multiple projection systems such as CAVEs (Cave Automatic Virtual Environment), and shutter glasses.

Fully immersive systems currently are based on the use of head-mounted display (HMD) technology. They give highest sense of immersion, although the image quality compared to previous technologies is worse.

Health issues are often raised in connection with immersive systems. Dizziness and lost of sense of orientation can be caused by the use of VE devices. Most commonly these problems are connected with the use of HMDs which are reported to physical, physiological and psychological problems. HMDs may also lead to posture problems because of the additional load on the body.

Augmented reality is a concept taken in contrast to virtual reality. Unlike virtual environments and their imaginary experiences, augmented reality uses the modern computer technologies to supplement the perceptions of the user of his actual surrounding. The basic factor for building augmented reality is not in different devices, but in the content of information provided to the user and its relation to the real world surroundings.

Care for Impaired Users

Another important trend in current technological developments is the greatest care for users with visual and other impairments. Thus people who were not able to use computers before join the computer society.

In the last years, we witness technological changes in the hardware every six months, and the appearance of novel devices. The history of computing never witnessed such plethora of devices with different technical capabilities and underlying concepts. The challenge to the users in this respect is to be open to new developments; form clear views on the necessity to master a new device and quickly learn to use it.

Software

The next level of interest in the computer culture is related to the employment of new hardware possibilities in the development of software.

In software development nowadays, several trends are influencing the development of systems, which are more attractive to the users:

- Personalisation of users in the virtual environments through use of avatars.
- Tele-presence applications, which contribute to form the sense of presence at another location.
- Building virtual communities.

Personalisation of Users

The virtual presentations of human users in the computer environment are called avatars. Human presentations of computer processes or programs are called agents. Both avatars and agents assist in forming the sense of social dimension in the human-computer interaction. Their use does not involve heavy bandwidth necessary for use of other technologies presenting the real user in the virtual environment, such as video conferencing and this explains their growing popularity.

Telepresence

Telepresence is defined as “the use of technology to establish a sense of shared *presence* or shared *space* among separated members of a group” [Buxton 1992]. Telepresence is a step ahead of video conferencing. It is of special importance due to the fact that it enriches communication with non-verbal aspects: gestures, eye contact, and spatial perception. Telepresence blurs the sense of reality of the user and moves him/her to another place.

Virtual Communities

Modern ICT offer a broad range of tools supporting communication and help people to be in contact with other people from all over the world. In the beginning, the communication tools were text-based and asynchronous, such as the e-mail and mailing lists. Later, instant messaging, multi user dungeons and video conferencing added other means and synchronised the communication. Peer-to-peer technology allows people in a community to exchange resources [Ross et al., 2004]. Virtual community is defined as a network of individuals/organisations using digital technology to create and use shared experience and knowledge.

Knowledge and *emotions* are the two typical centres around which communities are being built. In the first case, people collect and exchange opinions, learning materials, information how to do something or where to find particular content, etc. In the second case, communities gather to fulfil emotional or communication needs.

The Humans’ Basic Roles: Developers and Users

People immerse in the digital environment in two basic roles: developers and users. The users can be themselves creators, or passive consumers.

Fig. 1 presents the relationships between the different components of the digital culture.

In the digital eye centre (see 1) is the field of new digital technologies (I). Around it, from the centre towards the periphery, are differentiated its adjacent inheriting areas (II, III, IV). However, the links between them are neither linear nor hierarchical and these areas are overlapping and intersecting each other (see 2). Thus the image of the Digital Shell (see 3) is being formed. Surrounded by the amorphous area of the consumers (V), this metaphor gives an idea about the digital culture which conquers our society slowly, gradually, but undoubtedly.

The process is bi-directional (see (4)). The digital shell (A) develops and perfects its products around the needs, interests and expectations of its users (B). They play the role of the small grits which boost the development of mother-of-pearl – in this case of the research quests and the development of innovative technologies, while finally the pearl itself – i.e. the specific product, appears (see C). On this stage the new product has been delivered back to the user and influences his future interests. Thus, the developers of the digital products represented by the digital shell strongly influence the interests of the mass user. This two-fold process enhances the development of each tendency and also of the mistakes, if they had been done.

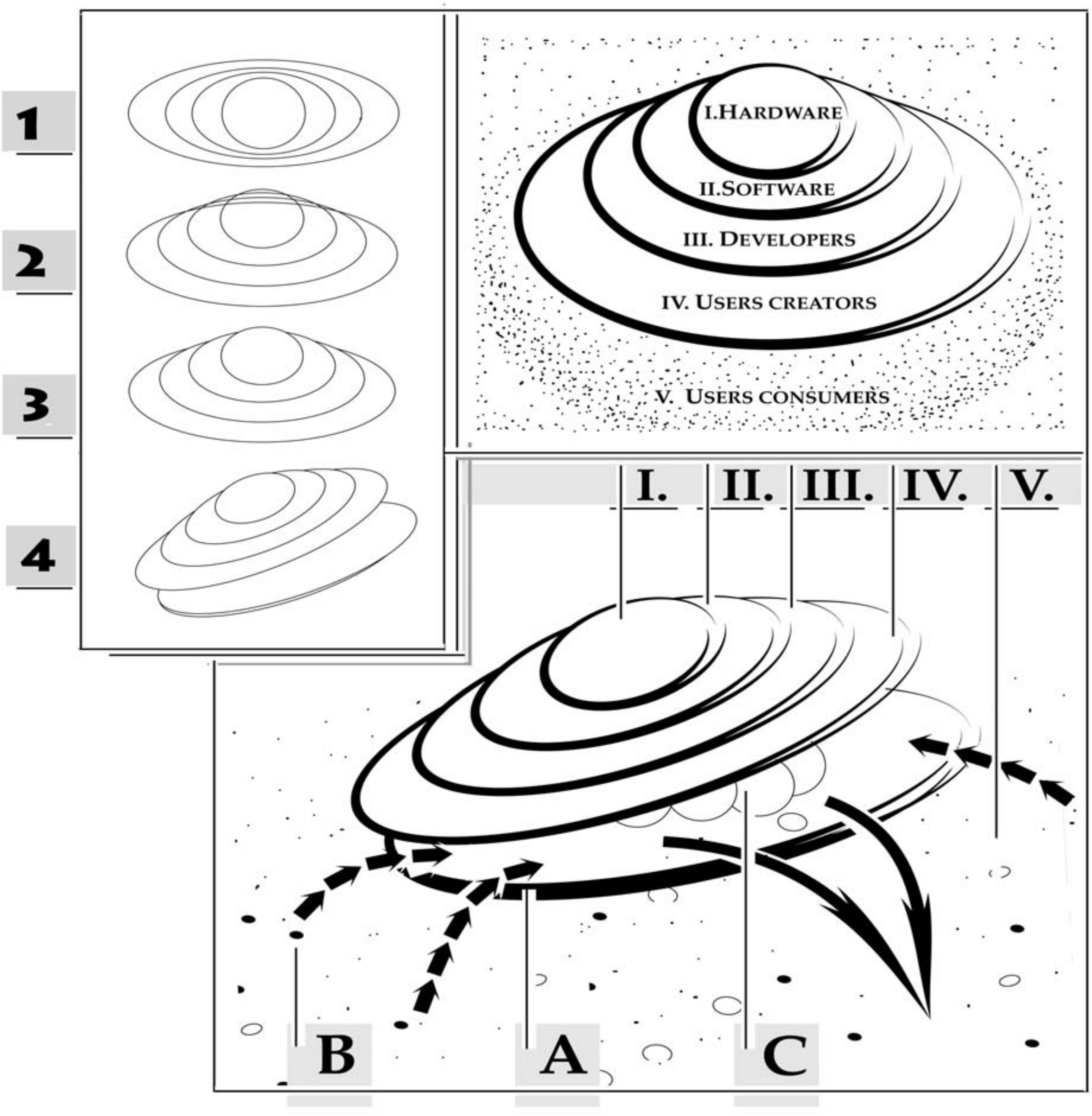


Fig. 1. The Digital Culture Ingredients

The digital shell should educate the users taking into account their interests, without allowing they to become leading ones. However, this is not easy to achieve in the market economy where the basic concern is the profit, which could lead to domination of the bad taste and low criteria due to the sporadic nature of the mass consumer.

Conclusion

Finally, we would like to summarise (see Table 1) the strengths-weaknesses opportunities and threats of the current virtual environments.

STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
Globality and dimension of the net: the new technologies can unify the world not only in the sense of information, but also in the sense of social issues. The problems being regional or private become global and shared.	The lack of order leads to chaos; many issues are not noticed which causes the 'voice in a desert' effect.	Dissemination of knowledge. Faster problem solving. Unifying the efforts of people from all over the world in problem solving. Building synergies in decisions making.	The effect of negative messages is multiplied.
Immense power of effect of the digital culture on the human consciousness.	Lack of order and control. Anonymity and lack of responsibility in the cases of misuse.	These technologies could contribute in the cases of global problems (ecological, terrorism, etc.) which solution requires behavioral change in humans. Raising of moral and ethical norms common to the mankind.	Manipulation of the consciousness, malicious use and propagation of false moral values. Deliberate or sporadic destroying of the moral ideals.
VE can be used to model and simulate almost everything one could imagine.	Immersion is still not 100% possible.	New possibilities are opened for training, education and entertainment.	The real life is shifted; asocial behavior can grow.
New educational opportunities: low price, mass coverage.	Knowledge obtained is incomplete (the practical experience and 'hidden' knowledge can be achieved only through personal face-to-face contact).	One can gain precious experience for situation, which are too dangerous to be learned in the real world (e.g., modelling of hazardous environments, calamities, space etc). Involvement of outstanding personalities in the learning process is easier in the VE than in the real world.	There is a danger to overestimate this type of education and to throw off the real practice, which could lead to abundance of low-level graduates.
Considerably greater possibilities for creative activities (professional, hobbies, and entertainment). Everyone could try himself and act as an artist, architect, musician, designer, animator, director, etc. without being formally trained to do this.	Diminishes the formal need for preceding professional training. Lack of natural censorship regarding works of art and artists, compared to classical arts. Every work despite its artistic value is made public.	Stimulation and development of the creative nature with appropriate creative games.	Invasion of unprofessionalism in the art. Building false creative self-confidence. Decay of the esthetical criteria and expectation of the user. Creation of clichés suffocates originality and creative thinking.

New possibilities for entertainment. Things impossible before become possible: immersion in activities and situations, wars, research missions, magic worlds, entry into impossible roles (warriors, magicians, emperors, Gods). Safety. Lack of fear of the consequences. Freedom and completeness of the experience.	Lack of connection between action and consequence in all personality aspects: physical, physiological, moral, social. Lack of risk, lack of fear of fatal errors.	The hypnotic power of the digital entertainment could be connected to the education in a way, which leads to improving its outcomes.	Weakening of the sensitivity and the instinct of self-preservation (making harm to self and others). Feeding the low human passions: aggression, power-loving, and cupidity. Lack of responsibility for one's actions in life. Lack of understanding of the importance and irreversibility of life events and developing of 'undo' behaviour.

Table 1. SWOT analysis: Digital Culture

This paper was written with the idea to attract the attention of IT professionals on the deep meaning of their work on the personal development and, respectively, on the society as a whole.

If we take as a starting point Nietzsche's thought "A person's maturity consists in having found again the seriousness one had as a child, at play."[29], we can say that current technologies are on the way to assist one to find the way back to his childhood by a number of external factors, such as hardware devices and interfaces. This cannot be claimed for seriousness, feeling, passion – and how often the developers or users themselves realise this?

It is unlikely that the ideological predecessors of the World Wide Web imagined what changes in human lives its appearance would actually cause. We should realize the responsibility. The mankind knows cases when its own creations turn against it.

We should not forget that digital technologies are not an end in themselves, but a tool for support and development in our lives. But they are also an element with an immense power, like the fire and nuclear energy. It depends only on us whether this element would serve us or will come out of control and how we will use it – for creation or for destruction.

Let we will be led in our digital creative acts by the humanity and the care for people; by the unique distinctive creative nature and potential of the human personality and the preservation of the human values in the society as a whole and not by the technology itself.

Bibliography

[Buxton, 1992] W.Buxton, *Telepresence: integrating shared task and person spaces*. Proceedings of Graphics Interface '92, 123-129., 1992. http://www.billbuxton.com/shared_space.html

[Kuhn 1962] T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, 1962.

[Negroponte, 1996] N. Negroponte, *Being digital,* First Vintage Book Edition, 1996.

[Polanyi 1962] M. Polanyi, *Personal Knowledge: Towards a Post-critical Philosophy.* Chicago: University of Chicago Press. 1962.

[Ross et al., 2003] S. Ross, M. Donnely, M. Dobreva, *New Technologies for the Cultural and Scientific Heritage Sector (Technology Watch Report 1),* ISBN 92-894-5275-7, European Commission, 2003, 196 pp.

[Ross et al., 2004] S. Ross, M. Donnely, M. Dobreva, *Emerging Technologies for the Cultural and Scientific Heritage Sector (Technology Watch Report 2),* ISBN 92-894-5276-5, European Commission, 2004, 214 pp.

[Sotirova, 2004] K. Sotirova. *Edutainment games – Homo Culturalis VS Homo Ludens*, In: Review of the National Digitisation Center, Belgrade, Serbia and Montenegro, Vol. 4, 2004, pp. 84–98.

Author information

Velimir Velev – Assistant Professor, National Academy for Theatre and Film Arts, Studentski grad bl. 58-V-204, Sofia-1734, Bulgaria; e-mail: velimir_velev@abv.bg

[29] Friedrich Wilhelm Nietzsche, Beyond Good and Evil (1885-86)

XML PRESENTATION OF DOCUMENTS USED FOR DATA EXCHANGE IN MANAGEMENT OF GENETIC RESOURCES

L. Yordanova, Vl. Dimitrov

***Abstract**: In the global strategy for preservation genetic resources of farm animals the implementation of information technology is of great importance. In this regards platform independent information tools and approaches for data exchange are needed in order to obtain aggregate values for regions and countries of spreading a separate breed. The current paper presents a XML based solution for data exchange in management genetic resources of farm animals' small populations. There are specific requirements to the exchanged documents that come from the goal of data analysis. Three main types of documents are distinguished and their XML formats are discussed. DTD and XML Schema for each type are suggested. Some examples of XML documents are given also.*

***Keywords**: XML, document's format, data exchange*

Introduction

The farm animals' genetic diversity is endangered and many breeds and lines extinct every year. World Watch List [Loftuse, 1992] with endangered breeds becomes longer as well as the list of lost breeds forever. The conservation of farm animals' genetic resources needs a sustainable management of small populations in each country and region. This is connected with the establishment of information systems for data collecting, maintaining individual records for the animals and relevant data processing for population analysis.

The management of genetic resources requires separate subsystems to exchange data or to send to a centre where aggregate values to be obtained for the region or country of keeping given farm animal population. It is possible separate system nodes to use different operating systems or database management systems. This could make difficult the data exchange between them. Therefore, they need of platform independent tools what do not restrict their communications. The implementation of XML standard could be a successful approach nevertheless the target area is very complicated and many options are possible in definition of used documents.

The current work is a part of an environment for developing information system for managing small population of farm animals. The first implementation of XML standard in the environment is connected with definition of a XML format for database model and creating implementation tools for its utilising [Yordanova, 2003].

The subject of current paper is to suggest XML formats for determined main types of documents used in the management of genetic resources. The analysis of all used documents restricts the discussion to three types:

- Documents connected with data streams
- Documents for data exchange in population analysis
- Documents for data exchange in other kinds of data analysis.

XML format of an auxiliary data stream document

The main type of documents exchanged in management of genetic resource is connected with data streams populating the database. A data stream is a document containing records of a same format. They could be repeated records for one animal or one record per each animal in a group. Such documents contain variety of concrete data elements and that is why their representation with a generic structure is difficult without high degree of abstraction. What we can do is to reach common XML format for the description of any data stream. If we ignore the concrete contain of the documents the result could be a very simple document tree with a root element ***stream*** and its descendent ***dataelement***. The suggested set of elements, even a simple one, will be enough for representation the structure of any document of a data stream.

The DTD of an auxiliary data stream is given in the listing 1. The root element ***stream*** is considered with an attribute for its name. The element ***dataelement*** would be well characterized with set of attributes ***name***, ***type*** and ***description***. This element could be at least once in a separate document of such type. Although it is impossible to have only one element in a document on practical reasons “once” could be accepted conditionally.

Listing 1. The DTD of XML format of a data stream

```
<DOCTYPE stream [
<!ELEMENT stream(dataelement+)>
<!ATTLIST stream                         Name          CDATA        #REQUIRED>
<!ELEMENT dataelement                    EMPTY>
<!ATTLIST dataelement                    Name          CDATA        #REQUIRED
                                         Type          (CHAR|HUGEINT|BIGINT|SMALLINT|
DATE|TIME|TIMESTAMP|SMALLFLOAT|BIGFLOAT|BOOL)          "CHAR"
                                         description   CDATA        #REQUIRED> ]>
```

In the XML Schema of a data stream the elements ***stream*** and ***dataelement*** are defined as Complex type and the attribute ***type*** has Simple type with enumerated values.

After definition of above XML format for the description of a data stream we must discuss the way of its usage. One possibility is such XML file to be attached to the document connected with the data stream in order to describe its structure. Then the application programs of the system could use it as a dictionary for data within the stream. They also could generate a set of commands inserting the data into the database. This seems to be a generic solution applicable to all possible data streams with different structure.

As a second possibility we consider the conversion of the data stream documents to the XML format that must be a solution of a separate information system. A separate XML format reflecting the structure of a given document could be developed and implemented there.

An example of a XML file containing a description of a data stream is given in listing 2. This one describes the data stream named Semen from the information system “Cryo” [Groeneveld, 2002].

Listing 2. A XML file with description of a data stream

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE stream SYSTEM "file:///home/lina/teza/potok.dtd">
<stream name="Semen">
    <dataelement description="The external id for the bull"
                 name="ext_animal"/>
    <dataelement description="The external id for the reporting unit"
                 name="unit"/>
    <dataelement description="The date of semen delivery"
                 name="delivery_dt" type="DATE"/>
    <dataelement description="The date of semen production"
                 name="production_dt" type="DATE"/>
    <dataelement description="The number of doses total"
                 name="no_doses" type="SMALLINT"/>
    <dataelement description="The number of straws per dose"
                 name="no_straws" type="SMALLINT"/>
    <dataelement description="The type of straws"
                 name="type_straws"/>
    <dataelement description="Quality mark- motility post freezing %"
                 name="mot_post" type="SMALLINT" />
    <dataelement description="Quality mark- after collection %"
                 name="mot_after" type="SMALLINT" />
    <dataelement description="Semen certificate identification"
                 name="certificate_id" type="SMALLINT"/>
</stream>
```

The XML file from the listing 2 is created and validated through defined DTD with XML editor what could be done also by any program application working with XML standard.

The documents of a data stream could be obtained on various approaches. Nevertheless which approach is used the documents of data streams have the logical structure defined above. The defined XML format of their description can be used in document exchange. In consequence data from corresponding XML document of a data stream could be inserted into the database via middleware software.

There are practical cases in management of genetic resources with manual filling in paper documents and than converting in any electronic form. The most common situation is conversion to a comma separated value format. Technical tools produce files of the same format often. It is possible also that the concrete document to be converted to its own XML format.

The documents of data streams are connected in general with the GUI forms for inserting and manipulating data [Yordanova, 2000]. The GUI forms in the environment are created according to them. The description of a data stream does not contain the actions concerning the database elements in order to insert data from the stream.

We should consider that the data streams are mainly connected with primary data collecting. It is very seldom the data they contain to be retrieved from another database but if this is the case then the approaches from next chapters are applicable.

XML formats of documents for population analysis

The data exchange in management of genetic resources covers different groups of data depending on the purpose of their analysis.

The population analysis needs of data about the animal origin. Such analysis requires obtaining of individual inbreeding coefficients, effective population size and other genetic parameters that the manager of breed conservation program could choose. Then the data exchange between the center and peripheral nodes must include: the identification of the animal, the identifications of its parents, gender and birth date or birth year. This set of data is a minimum, enough for calculation the genetic parameters for population analysis.

We define the XML format of document containing data for animal origin and its individual identification via DTD (listing 3) and XML Schema.

Listing 3. The DTD of a document for data exchange in population analysis

```
<DOCTYPE pedigree [
<!ELEMENT pedigree(animal+)>
<!ATTLIST pedigree                  name     CDATA     #REQUIRED>
<!ELEMENT animal(birthdt, sire, dam)>
<!ATTLIST animal                    ext_id   CDATA     #REQUIRED
                                    unit     CDATA     #REQUIRED
                                    gender   (F | M)   #REQUIRED>
<!ELEMENT birthdt(#PCDATA)>
<!ELEMENT sire(#PCDATA)>
<!ATTLIST sire                      ext_id   CDATA     #REQUIRED
                                    unit     CDATA     #REQUIRED>
<!ELEMENT dam(#PCDATA)>
<!ATTLIST dam                       ext_id   CDATA     #REQUIRED
                                    unit     CDATA     #REQUIRED> ]>
```

The root element is called ***pedigree*** and its sub element is ***animal***. The sub element is defined to be at least once in the document. It has attributes ***ext_id***, ***unit*** and ***gender*** and sub elements ***birthdt***, ***sire*** and ***dam***. The elements ***sire*** and ***dam*** should have the same attributes like the element ***animal***. In all cases the attribute ***ext_id*** means an external identification of an animal depending on the unit that reports the animal. Here it is not convenient to use for ***ext_id*** type ID, because in common case it is not unique. Two units can report two animals with the same identification. That requires the couple of elements (***ext_id, unit***) to be unique.

The element or attribute connected with information about the breed to which the animal belongs is not included. This is done because the population analysis supposes that collecting data concerns animals from the same breed. If it is necessary one could mark the breed into the name of the document or to add an element breed. For the example given here the breed name is the content of the attribute ***name*** of the root element (listing 4). The example document is for population analysis according the XML definition explained so far. The data is for two family couples. Four animals (the progeny of the families) are included. The document is checked for validation with corresponding XML schemas through program applications that use DTD or XML Schema.

Listing 4. An example for data exchange in population analysis

```
<?xml version="1.0" encoding="UTF-8"?>
 <!DOCTYPE pedigre SYSTEM "file:///home/lina/teza/pedigre.dtd">
 <pedigre name="minipigs">
   <animal ext_id="1677" gender="F" unit="12">
            <birthdt>23.12.1998</birthdt>
            <sire ext_id="3456" unit="12">04.04.1997</sire>
            <dam ext_id="2345" unit="12">12.03.1996</dam> </animal>
   <animal ext_id="1698" gender="F" unit="12">
            <birthdt>23.12.1998</birthdt>
            <sire ext_id="3456" unit="12">04.04.1997</sire>
            <dam ext_id="2345" unit="12">12.03.1996</dam> </animal>
   <animal ext_id="1701" gender="M" unit="12">
            <birthdt>3.11.1998</birthdt>
            <sire ext_id="5003" unit="12">12.08.1996</sire>
            <dam ext_id="4312" unit="12">12.03.1996</dam> </animal>
   <animal ext_id="1702" gender="F" unit="12">
            <birthdt>3.11.1998</birthdt>
            <sire ext_id="5003" unit="12">12.08.1996</sire>
            <dam ext_id="4312" unit="12">12.03.1996</dam> </animal>
   </pedigre>
```

XML formats of documents for other kinds of analysis

The common structure for documents exchanged in the management of genetic resources with the goal to perform other kinds of analysis contains mandatory animal identification and its one or multytrait measurements. A possible generic XML scheme of such document is given in the listing 5 with DTD. The root element named ***data*** is consisted by at least one sub element ***animal***. It has attributes ***ext_id*** and ***unit***s as well as one sub element trait meet more than once. The element ***trait*** connects any investigated trait to an animal. It is possible a document to have data about more traits that complete a process. One animal could be measured many times for a trait. That is why it is appropriate to have a sub element ***measurement*** repeated many times. The measurement is characterized with attributes or sub elements ***date***, ***value*** and ***type***. The measurements type requires from the application programs to maintain with external coding for different types of measurements.

The animal identification is given with attributes ***ext_id*** and ***unit***, which means maintaining external identification for both objects, ***animal*** and ***unit***. This requires from the software to obtain a new or to retrieve existing internal identification from the database. The last one is used according the system supporting unique identification for the animals. About reporting unit it is most possible to have only second situation.

Listing 5. The DTD of a document for data exchange in other kinds of analysis

```
<DOCTYPE data [
<!ELEMENT data(animal+)>
<!ELEMENT animal(trait+)>
<!ATTLIST animal              ext_id    CDATA     #REQUIRED
                              Unit      CDATA     #REQUIRED>
<!ELEMENT trait(measurement+)>
```

```
<!ELEMENT measurement EMPTY>
<!ATTLIST measurement          Date    CDATA    #REQUIRED
                               Value   CDATA    #REQUIRED
                               Type    CDATA    #REQUIRED> ]>
```

A XML document obtained on the defined format is given in the listing 6. The document is validated according the XML format definition. It contains weight measurements of two animals for tree months.

Listing 6. An example of XML document with data for the trait weight

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE data SYSTEM "file:///home/lina/teza/data.dtd">
<data>
   <animal ext_id="6768" unit="11">
        <trait name="weight">
             <measurment date="12.03.2000" type="bdw" value="0.3"/>
             <measurment date="12.05.2000" type="wdw" value="0.6"/>
             <measurment date="14.06.2000" type="rw" value="1.9"/>
             <measurment date="13.07.2000" type="us" value="2.4"/>
        </trait>
   </animal>
   <animal ext_id="4546" unit="11">
        <trait name="weight">
             <measurment date="14.04.2000" type="bdw" value="0.2"/>
             <measurment date="15.06.2000" type="wdw" value="1.7"/>
             <measurment date="14.07.2000" type="us" value="2.6"/>
        </trait>
   </animal>
</data>
```

For the documents of data streams and for the data exchange it is preferable to get data automatically. Retrieving data from the database is common for the documents for data exchange. The cases where data is generated from computer systems with measure tools automatically are very seldom, especially in small populations. Then generated files are usually in CSV format. Their transformation to the XML format is not a difficult task. Better solution is the XML format defined for a data stream to be used as a meta form of such CSV file. The middleware software can operate with data according the description of the document structure in XML format.

Middleware software for data exchange

In current work the XML documents are used for transfer data after it is generated or parse by dynamic program application. The main processes discussed here are:

- Parsing and retrieving data from a XML document in order to be inserted into the database
- Creating a XML document retrieving data from the database.

Both program codes use the XML formats of the documents in data exchange defined above.

The program code for inserting data into the database from a XML document

Let the XML document that is going to be parsed and analyzed in order to insert data into the database has the structure from listing 3. Let the target database has the conceptual database scheme [Yordanova, 2000]. Then the algorithm for parsing the documents and inserting the data into the database includes the next steps:

```
Begin                     (#begin
DBI connection            (# Connection to the database
Objects and variables     (# Declaration of objects and variable
SQL statements            (# Definition of SQL statements -
Foreach $row              (# For each animal retrieving of:
   Sire/ID, sire/unit        (# db_sire from TRANSFER(ext_id, unit)
   Dam/ID, dam/unit          (# db_dam from TRANSFER(ext_id, unit)
```

```
    Animal/unit                         (# db_unit from UNIT via unit
    get_next_val(sequence_name)         (# new db_animal identification
    INSERT into ANIMAL, TRANSFER        (# Inserts in TRANSFER and ANIMAL
  end foreach                         (# End of the cycle
  db disconnect                       (# Disconnect the database
  End                                 (#End
```

If a parent does not have internal database identification then a new one has to be obtained in the relation TRANSFER and recorded into the relation ANIMAL. The inserts will not be done if there is another animal with the same values of (ext_id, unit). The released program is a Perl code and uses the module XML::XPath of Matt Sergeant that implements the XPath standard and allows fast search and parsing elements of XML document via tree of document's nodes.

The program code for retrieving data from database

The function of the code is connected with the XML format for population analysis (listing 3). The algorithm is separated to two main steps:

1. Getting via Query a set of tuples, containing the external identification of all active animals and their parents as well as their reporting units, birth dates and gender.
2. Recording data from the tuples in XML document elements.

This code uses Perl&XML module XML::Writer that allows creating of XML document via defined objects.

Conclusion

The usage of XML standard makes the data exchange in management genetic resources much more flexible and platform independent. There are a lot of program applications that work with many operating systems and could facilitate implementation of defined XML formats of documents for data exchange. The user could create the XML documents containing the description of data streams using: 1) XML editors that apply XML declarations DTD or XML Schema; 2) program applications that includes processing and validating XML documents through schema.

The other XML documents for data exchange in all kinds of data analysis for small populations could be created and used via briefly presented here program codes.

The defined XML formats could be extended and could become a base language for data exchange in management of genetic resources.

Bibliography

[Groeneveld, 2002] E. Groeneveld, L.Yordanova and S.J. Hiemstra. An Adaptable Management Sopport System for National Genebank, 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpelier, France, Session 26, Management of genetic diversity, 513-516, 2002.

[Loftuse, 1992] R. Loftuse and B. Scherf, World Watch List for Domestic Animal Diversity, FAO Rome, 1992.

[Yordanova, 2000] L. Yordanova, E. Groeneveld. The implementation Strategy of Information System with Existing Data Streams, Vortrastagung der Deutschen Gesellschaft fuer Zuchtungkunde e. V. (DGfZ) u. der Gesellschaft fur Tierzucht Wissenschaft (GfT), A28, 2000.

[Yordanova, 2003] L. Yordanova, E. Groeneveld, Vl. Dimitrov. A XML Approach for Support DB Modeling. In: Proceedings of the Thirty Second Spring Conference of the Union of Bulgarian Mathematicians, 297-302, 2003.

Author information

Lina Yordanova – Thracian University, Department of Informatics, Mathematics and Physics, Stara Zagora, 6000, Studentsko gradtche, Bulgaria; e-mail: lina@uni-sz.bg

Vladimir Dimitrov – Sofia University, Faculty of Mathematics and Informatics, Sofia, Bulgaria; e-mail: cht@fmi.uni-sofia.bg

ALGORITHM BIDIMS FOR AUTOMATED SYSTEMATIZATION OF DATA ARRAY. CASE STUDY: REDISCOVERING MENDELEEV'S PERIODIC TABLE OF CHEMICAL ELEMENTS.

Andrey N. Zagoruiko, Nickolay G. Zagoruiko

Abstract: *The method (algorithm BIDIMS) of multivariate objects display to bidimensional structure in which the sum of differences of objects properties and their nearest neighbors is minimal is being described. The basic regularities on the set of objects at this ordering become evident. Besides, such structures (tables) have high inductive opportunities: many latent properties of objects may be predicted on their coordinates in this table. Opportunities of a method are illustrated on an example of bidimentional ordering of chemical elements. The table received in result practically coincides with the periodic Mendeleev table.*

Keywords: *bidimensional structure, data mining, ordering, prediction, approximation.*

Introduction

One of Data Mining purposes consists in reduction of the available data to such kind at which the person easily perceives the basic contents of the analyzed information. The information advanced in such a way becomes a little more substantial for the person. The specified purpose is achieved by different means. The important role plays machine graphics, allowing the person to perceive the information through the powerful visual analyzer. Various ways of ordering at which one-dimensional or bi-directional data file will be coordinated well with the simple concepts (models) have a wide circulation. One-dimensional line of smoothly growing or decreasing numerical values is evident, for example. Even more information is contained in bi-directional tables with monotonic change of data on the first and second coordinate. It is not surprising that many fundamental laws of a nature - the law of the Ohm, Newton, Mendel etc. are well illustrated by bi-directional data tables.

D.I. Mendeleev, studying dependencies between various properties of chemical elements, relied on many results of the predecessors. In particular, the grouping of some elements on a generality of their chemical properties was known, on similarity of nuclear weights etc. Mendeleev put the task before itself to make such bi-dimensional ordering of all 63 elements known at that time at which the neighbouring elements of the table would be similar to each other on maximum big set chemical and physical properties. Such their arrangement will be coordinated with the concept of local smoothness, which provides easy perception of the general laws for the table.

The use of concepts beforehand prepared or the models is useful not only for an explanation of the data analysis result, but also for the process of this result receiving. The rich history of scientific discoveries speaks about it, in many of which traces of attempts of such type are obviously visible: " - It seems these objects have a S-structure. And what if to try use such a model? And what if to order the objects by such rule?"

Let's illustrate the utility of the initial empirical data association with the simple concepts during revealing the laws latent in data. We'll do it on an example of automatic rediscovering of the periodic law of chemical elements. Such attempt was already done by us [1], but in a little bit idealized conditions. In particular, true nuclear weights and valences on hydrogen, instead of those known to Mendeleev were used; the program was adjusted to the fixed number of properties etc. In the given work those data and knowledge, which D.I. Mendeleev had at his disposal during the creation of the periodic law of chemical elements, are used.

Description of systematization algorithm

Let us imagine that we have some data array, consisting of *n* elements *A*, each of them characterized by a set of *k* properties. In the current formulation the task is to systematize this array in a form of a two-dimensional table with internal uniformity in changing of element properties in both dimensions.

Let us also propose that we have some «embryo» of the table, i.e. credible group of small amount of elements, constructed on the base of researcher intuition. Having such a group it is possible to try to predict the properties of neighbouring elements. Assuming the uniformity of elements properties changing inside the «embryo» in both dimensions, this procedure may be performed using linear approximations. Possible types of such approximations are given below:

Internal approximation may be applied if it is necessary to predict the properties of the element, situated between two elements provided that their positions in the table are already known. In this case if known elements are situated in the table at coordinates, for example, *(i+1,j)* and *(i-1,j)*, then the value of *m-th* property in position (i,j) P_{ij}^m may be predicted using the equation

$$P_{ij}^m = \frac{P_{i-1j}^m + P_{i+1j}^m}{2} \quad (1)$$

Similar estimations may be obtained as well for combinations *(i+1,j+1)* и *(i-1,j-1), (i+1,j-1)* и *(i-1,j+1), (i,j+1)* и *(i,j-1)*, i.e. via the horizontal, vertical and two diagonals (total – 4 variants).

External approximation is used when it is necessary to construct the forecast for the table cell adjacent to the pair of situated elements with known positions. For example, if known elements are placed in the cells with coordinates *(i-2,j)* and *(i-1,j)*, then value of P_{ij}^m may be predicted as follows:

$$P_{ij}^m = 2P_{i-1j}^m - P_{i-2j}^m \quad (2)$$

Here the predictions also can be made via horizontals, verticals and diagonals (total – 8 variants).

Corner approximation is applied when known elements are placed in the table in form of "corner", for example, in the cells with coordinates *(i+1,j), (i,j+1)* and *(i+1,j+1)*. In this case it is necessary to use the equation

$$P_{ij}^m = P_{i+1j}^m + P_{ij+1}^m - P_{i+1j+1}^m \quad (3)$$

Four variants of "corner" positions are possible. The final prediction of the property P_{ij}^m is defined as averaged value of all forecasts according to equations, which total number may reach 16.

The next step of the procedure is selection of optimal "pretending" element from the set of remaining elements, which are not positioned in the table. The running through of all remaining elements in relation to every empty cell of the table is made with definition of the positioning quality of every element/cell combination. The modulus of deviation between predicted and real element property values may be used as quality criterion:

$$X_{ij}^h = \sum_{m=1}^{k} abs(P_{ij}^m - R_h^m) \quad (4)$$

where X_{ij}^h - quality criterion of *h-th* element in the cell with coordinates *(i,j),* R_h^m - real value of *m-th* property for this element. After running of every elements versus every table cell, the table is filled with only one element, which at all set of *h, i* and *j* is characterized with minimum value of X_{ij}^h. This procedure is repeated until completion of positioning of all initial elements.

The problem may be complicated by two factors. Firstly, range of parameter values may be significantly different for different properties, resulting in different contribution of each property to the value of X_{ij}^h criterion and, therefore, leading to their «inequality of rights». Secondly, it is not evident that every element is described by a full set of properties, so property array actually may include missing values. Correspondingly, the reversed situation is possible as well, when the definite element property is really present, but cannot be predicted, because it is not present in the property sets of elements used for prediction.

First complication is easily solved by normalization of data for each of properties. In the second case it becomes necessary to define for each cell and each element how many properties area really predicted and then normalize criterion X_{ij}^h value as follows:

$$X_{ij}^{h} = \frac{\sum_{m=1}^{k} abs(P_{ij}^{m} - R_{h}^{m})}{Z_{ij}^{h}} \tag{5}$$

where Z_{ij}^{h} - number of coincidence «successfully predicted property» / «presence of that property in the element property set» under attempt to place the *h-th* element in the cell with coordinates *(i,j)*. Moreover, as it was demonstrated by test calculations, to improve systematization quality it is better to give preference to elements with higher value of Z_{ij}^{h}, as more reliably determined. Such preference may be realized by different ways, but in this work we used empirical method, based on application of Z_{ij}^{h}, raised to a power higher than one. Particularly, the optimal order value was found to be 3, i.e equation (5) was transformed into:

$$X_{ij}^{h} = \frac{\sum_{m=1}^{k} abs(P_{ij}^{m} - R_{h}^{m})}{(Z_{ij}^{h})^{3}} \tag{6}$$

Systematization of a full set of chemical elements

As a case study we used a complete set of chemical elements (see Fig.1), e.i. the essential aim of the work was reopening of periodic law, discovered by Dmitry Mendeleev in 1869.

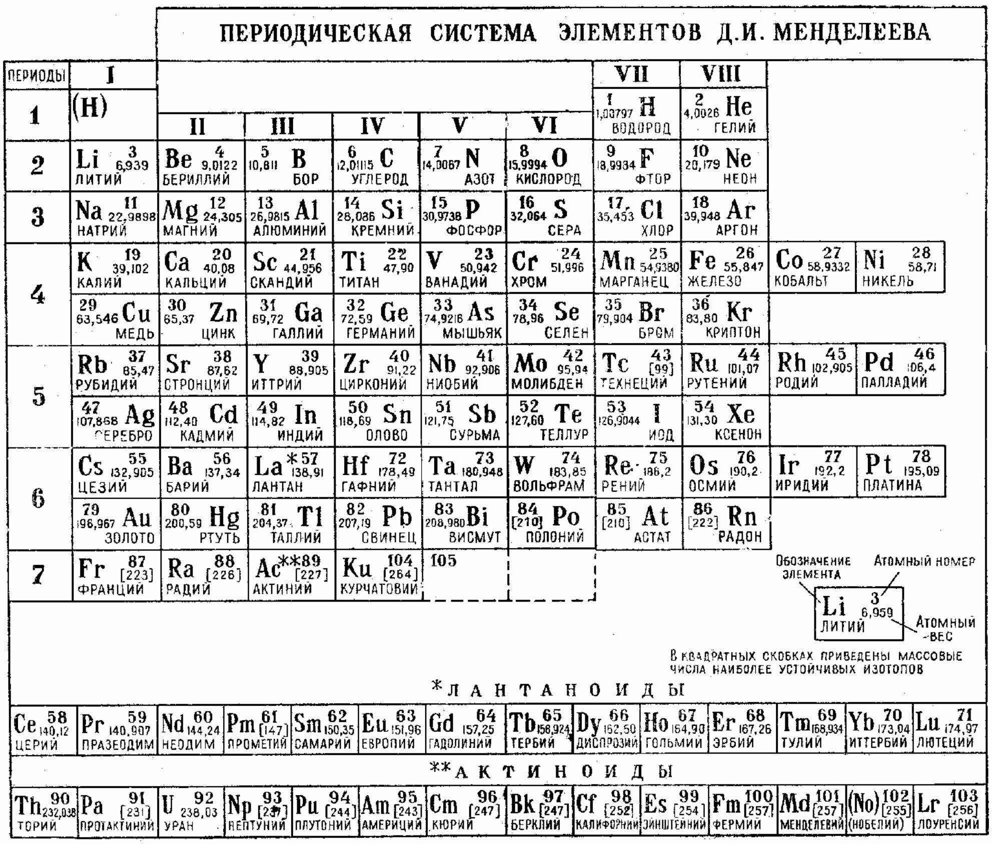


Fig.1. Short-period table of elements

At the first stage we performed test calculations with application of full set of chemical elements, known at the current moment. In case when the set of three basic properties (atomic mass, group number and period number) the described algorithm provided fast and correct solution. Of course, application of group and period numbers was equivalent to inclusion of already known correct solution into initial data. Therefore, this variant was for software testing purposes only.

On the second stage we used a set of three basic properties, that were known to Mendeleev and which were used by him in his work on periodic law construction: atomic mass, oxygen and hydrogen valence. As "embryos" we used intuitive combinations which look, nevertheless, quite obvious from chemical point of view, such as:

Na			O	F
K	Ca	or	S	Cl

In this case it was possible to successfully construct the "framework" of the table, where periodicity and uniformity of properties changing are evident, namely – 2nd and 3rd periods. The following table filling met significant complications, mainly for transitional element and, especially, for triads Fe-Co-Ni, Ru-Rh-Pd, Os-Ir-Pt and inert gases. Correct positioning of lanthanides and actinides was found to be completely impossible.

At the same time some interesting regularities were discovered. It was found that even one erroneous positioning of an element leads to the chain of further errors, the sooner the error is made the more significant distortions are contributed to the final result. It was also detected that table construction quality (quite logically) depends very much upon the choice of initial "embryo".

Explanations of all these problems are rather simple. Periodicity and uniformity of changing of atomic mass looks evident (Fig.2), at least if will not consider natural mass gap in the area of lanthanides placement (this gap is absent in a long-period table). At the same time the picture for oxygen and hydrogen valences is much more complicated (Figs. 3 and 4).

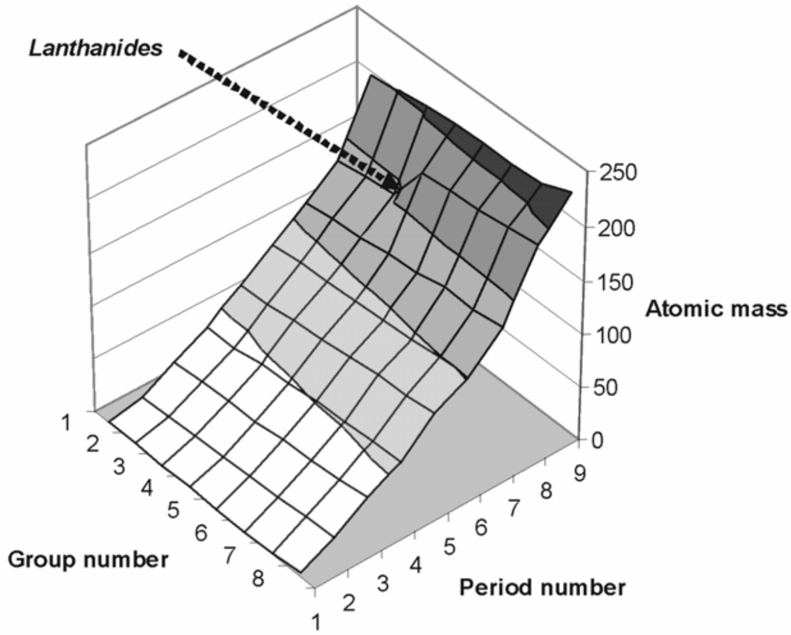


Fig.2. Changing of atomic mass of elements in groups and periods of "short" periodic table.

Good periodicity is observed for hydrogen valence (Fig.4), but this data is present for less than a half of elements. Furthermore, continuity of data is present in 2nd and 3rd periods only, and in higher periods the periodicity is broken by transitional elements (i.e. is repeated "a string after").

Majority of oxygen valence data (Fig.3) is fit into irreproachable flat plane, but with significant anomalies at the table periphery, notably:

- decrease of observed valence in triads Fe (6+) - Co(3+) - Ni(2+), Ru(8+) – Rh (4+) – Pd(2+), Os (8+) – Ir(4+) – Pt(4+);
- zero valence for inert gases (except Xe(8+) and Kr(2+));
- high valences of copper (2+) and gold (3+) instead of expected (1+).

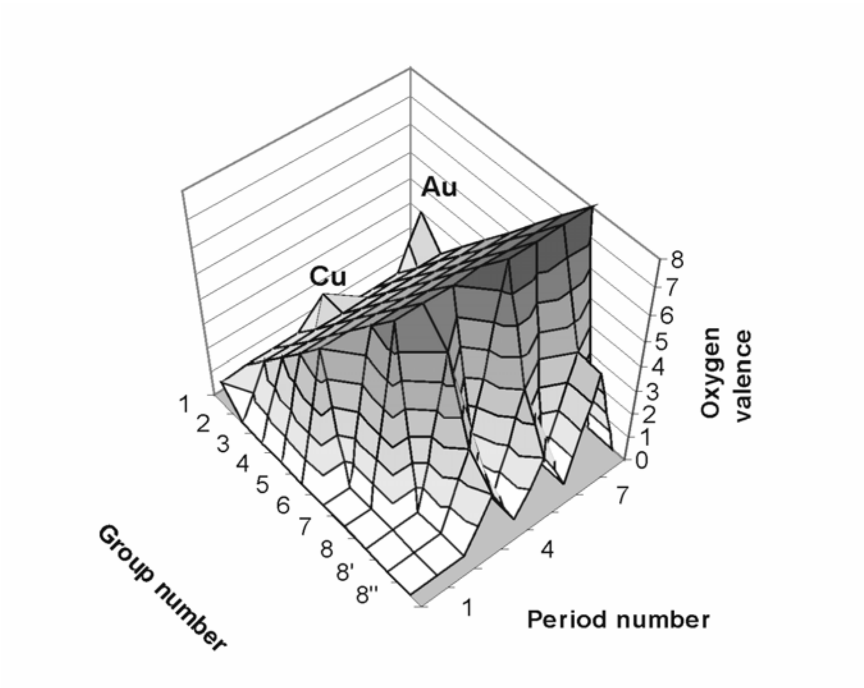


Fig.3. Changing of oxygen valence.

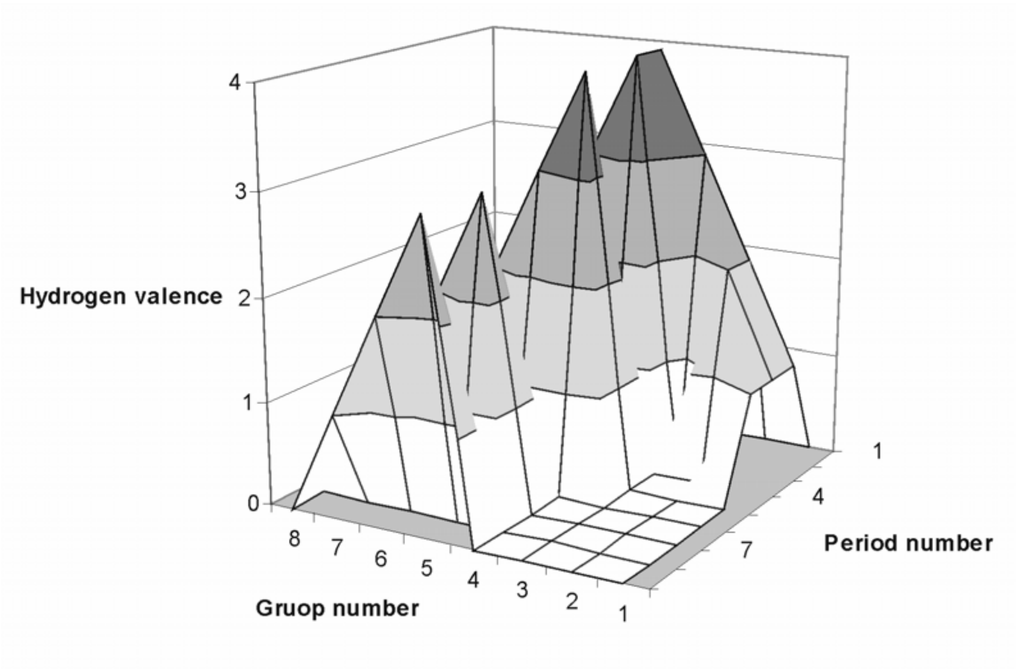


Fig..4. Changing of hydrogen valence.

In general it may be stated that in relation to mentioned properties the data array in Mendeleev's table is not uniform, as it was proposed at the stage of problem formulation. Nevertheless, the attempts were made to "smooth" these nonuniformities by application of greater number of properties in element descriptions. The properties that were definitely known to Mendeleev during his work on Periodic Law (densities, melting and boiling temperatures for elements and their oxides and chlorides; acid/base properties of oxides etc) were chosen to expand data array.

Surprisingly, this attempt was even less successful. The reason for this fault may be demonstrated on the base of element density changing (Fig.5).

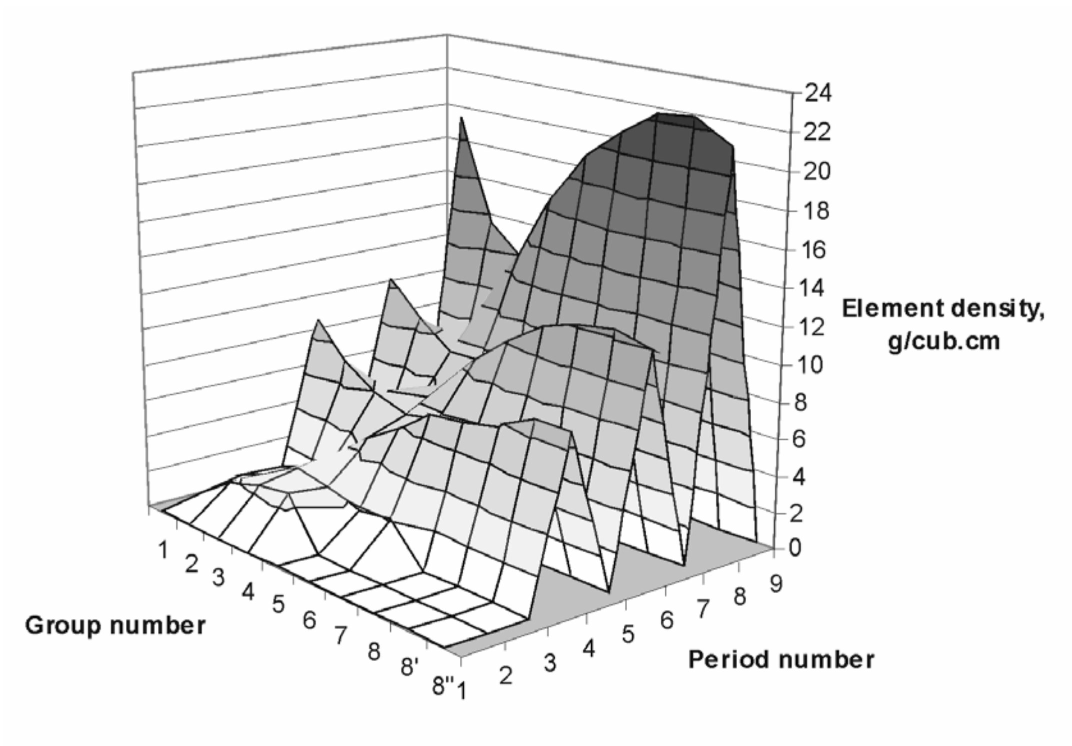


Fig.5. Changing of density of elements in groups and periods of "short" periodic table.

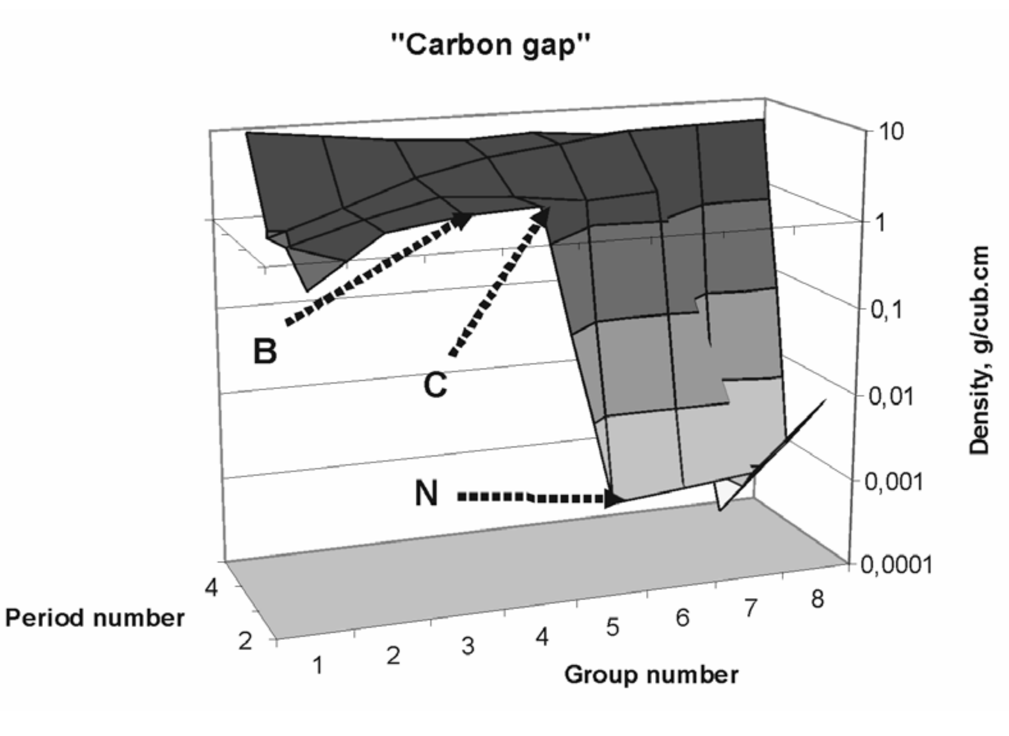


Fig.6. Elements density gap in the 2nd period.

Fig.5 shows strong nonlinearity and nonuniformity of plotted surface. Notably, such oscillations are typical not for densities only, but for all other mentioned properties as well. In principle, if we will switch to long-period table, then these oscillations will become smoother, but such switching is incorrect, because it will require choice of elements from 4th and higher periods for formation of the "embryo", what looks unobvious. Construction of table from more reasonable "embryo" elements of 2nd and 3rd periods will in any case lead to short-period type of the table.

Moreover, there also exists the problem that cannot be resolved even by transfer to long-period table. Let us consider the fragment of Fig.5 related to 2nd period. Here (Fig.6) it is seen that between starting elements in this period (Li, Be, B, C) and further sequence (starting from N) there is a gap in densities (by a few orders of magnitude – values in vertical axis are given in logarithmic scale). Similar anomaly behaviour is also observed for other physical properties. Gap position may shift from C/N to C/B zone (e.g. for melting and boiling temperatures of oxides), but is always connected with carbon. Existence of such anomaly ("carbon gap") finally leads to impossibility to construct correctly even the 2nd period of the table and to completely absurd construction of following periods.

Therefore, expansion of data array by introduction of additional physics-chemical properties only decreases the quality of systematization process. Of course, it is possible to try using modern elements characteristics (atomic electron configuration, ionization potential, electro negativity etc.), but it looks incorrect, because this information was not available to Dmitry Mendeleev, and, besides, in this case the data array will artificially contain correct solution, thus "killing intrigue" of this study.

Systematization of "Mendeleev's" set of elements

Though the systematization of complete set of elements was unsuccessful, in this pert of the study we tried to systemize the elements set that was known to Mendeleev during his work. This set is different from complete one in following:

- inert gases are completely absent (they were discovered later);
- majority of lanthanides and actinides, as well as heavy elements (heavier this bismuth) are absent;
- few elements from middle periods are also undiscovered (Sc, Ga, Ge).

Furthermore, for some elements Mendeleev has doubtful and incorrect data for atomic weights and valences. Due to aforementioned reasons here we used only data on atomic mass and valences. It should be noted that Mendeleev also used this information as basic in table construction process.

Surprisingly, it was found that in this case systematization is much more simple than that for complete set of elements. First of all, it is explained by absence of inert gases, which actually are placed in the table quite illogically (none of them, except xenon, demonstrate 8+ oxygen valence, which is predicted for this group). Additional advantage is absence of lanthanides, because their position in short-period table also looks quite unusual.

In this case we've managed to reproduce the major part of the table, but here as well we met the effect of "wrong" behaviour in triads of transitional elements. For example, cobalt (quite logically) was "trying" to fill the cell of absent gallium due to coincidence of maximum oxygen valence (3+) with relatively low error in prediction of atomic mass. From chemical point of view it is absolutely evident that cobalt is an analogue of iron and nickel, but not aluminum (as gallium), but such "chemical" understanding cannot be described in data array within existing data structure. Anyway, incorrect cobalt positioning led to distortions in further construction of the table. The same may be told about triads of noble metals.

Therefore, we've an attempts to modify initial data array, based on exclusion of the most "odious" elements, particularly:

- all elements from transitional triads, except first ones (Fe, Ru, Os), were excluded;
- present lanthanides were excluded (except La and Ce only).

Moreover, for copper and gold the basic oxygen valence 1+ was stated, though their actual maximum valences are higher (2+ for Cu and 3+ for Au). It was done to reveal the fact that Cu and Au are analogues of silver.

The result of systematization in this case is shown in Fig.7.

Period	Group							
							H	
	Li	Be	B	C	N	O	F	
	Na	Mg	Al	Si	P	S	Cl	
	K	Ca		Ti	V	Cr	Mn	Fe
	Cu	Zn			As	Se	Br	
	Rb	Sr	Y	Zr	Nb	Mo		Ru
	Ag	Cd	In	Sn	Sb	Te	I	
	Cs	Ba	La	*Ce*	Ta	W		Os
	Au	Hg	Tl	Pb	Bi	*U*		
				Th				

Fig.7. Result of systematization of “Mendeleev’s” set of elements.

It is seen that systematization quality is quite high. Practically all elements are placed in cells, where they should be. Exclusion is made by uranium which actually should be situated in V-7 cell after thorium, but this error is not important (Th and U actually should be placed in separate subgroup of actinides, which is stipulated in this type of the table). Placement of thorium and cerium also does not look formally correct, but actually it is quite usual for them to demonstrate 4+ valence, what gives the ground to position them in the IV-th group of basic table. Such their dual behaviour is well known and is defined by objective specifics of their electronic structure, so such placement may be accepted as appropriate. It is curious, that D.I. Mendeleev the same as also our program, has placed in the initial kind of the table Thorium and Cerium in 4-th group. Moreover, in the same group he has placed and Lantan [2,3]. We shall note that our program has placed Lantan on a correct place in third group.

Prediction of undiscovered elements properties

Special attention should be paid to prediction abilities of the constructed table. As it is seen from Fig.7, after systematization few cells inside the table were left unfilled. These cells strongly correspond to existing elements, that were undiscovered at the time of Mendeleev’s study. To predict the properties of missing elements we used the described algorithm. In this case only obtained data values being inside the normalized range ([0,1]) were chosen.

Result of such prediction is quite impressive. First of all, 5 elements that must be positioned inside the table body were clearly shown (their positions in the table are shown at Fig.7 by crossed cells). Description of predicted values is given in Table 1.

Table.1

Position column/string	Atomic mass		Oxygen valence		Hydrogen valence		Real element
	forecast	fact	forecast	fact	forecast	fact	
3/4	43,90	44,95	3	3	-	-	Sc
4/5	69,00	72,59	4	4	4	4	Ge
3/5	65,20	69,72	3	3	-	-	Ga
7/6	101,10	98,91	7	7	1	-	Tc
7/8	176,80	186,20	7	7	-	-	Re

It is seen that coincidence between predicted and actual property values is quite good. Moreover, during analysis of predictions that were excluded from consideration, because the predicted values were found to be outside normalized range, we selected the group of similar elements, which formally should have been positioned in 8th group and have formal valences 8+ for oxygen and 0 for hydrogen (shown by shadowed cells at Fig.7). The reason of their exclusion was zero hydrogen valence, what was considered as inappropriate property value. Actually these predictions are strongly equivalent to the group of inert gases and good coincidence between predicted and actual properties is seen here as well (see Table.2). Furthermore, "wrong" prediction of zero hydrogen valence in this case achieves real physical sense – inert gases do not form hydrogen compounds in reality.

Table.2

Position column/string	Atomic mass		Oxygen valence		Real element
	forecast	fact	forecast	fact	
8/1	6,00	4,00	8	-	He
8/2	20,05	20,18	8	-	Ne
8/3	35,80	39,95	8	-	Ar
8/5	80,30	83,80	8	2	Kr
8/7	138,80	131,30	8	8	Xe

It is interesting that after selection of predictions no forecasts were made non-existent elements, i.e. the algorithm has made no attempts to fill empty cells of 1st period and cells to the right and to the left of the table body.

Conclusion

In general we may state that proposed algorithm BIDIMS (under definite assumptions and modifications) successfully managed to systemize "Mendeleev's" set of elements and, in fact, repeated the discovery of Periodic Law in a form, that was possible in Mendeleev's work period.

Performed study, nevertheless, is not diminishing Mendeleev's achievements in any extent. First of all, used assumptions and modifications were based on intuitive and forced decisions, having no formally strong grounds. In second, the basic decisive properties (atomic mass, valences) were chosen the same as ones used by Mendeleev. And the most important – the essence of genius Mendeleev's discovery is proposition, that existing element may be systemized in form of two-dimensional table. We used this proposition as acknowledged fact in our study.

Bibliography

1. A.N. Zagoruiko and N.G. Zagoruiko. Experiments on Rediscovering of Mendeleev's Low by Using Computer // Strukturnyi Analiz Simvol'nikh Posledovatel'nostei (Vychislitel'nye Sistemy 101, Ak. Nauk SSSR, Novosibirsk, 1984), p. 82-90. (In Russian).
2. A.A. Makarenia. D.I. Mendeleev and physicochemical sciences. Atomizdat. Moscow, 1972, 256 p. (In Russian).
3. http://chemlab.pc.maricopa.edu/periodic/foldedtable.html
4. N.G. Zagoruiko. Applied Methods of Data and Knowledge Mining. Ed. Inst. of Mathematics. Novosibirsk, 1999. 268 p. (In Russian).

Authors' Information

Andrey N. Zagoruiko – Institute of Catalysis of SD RAS, e-mail: zagor@catalysis.nsk.su

Nikolay G. Zagoruiko – Institute of Mathematics SD RAS, e-mail: zag@math.nsc.ru

PAPERS in RUSSIAN and BULGARIAN

ПРИМЕНЕНИЕ НЕКЛАССИЧЕСКИХ ЛОГИЧЕСКИХ МЕТОДОВ В ЗАДАЧАХ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ И ПОИСКА ЗАКОНОМЕРНОСТЕЙ В БАЗАХ ЗНАНИЙ*

В.С. Йоцов, В.С. Сгурев, А.В. Тимофеев, Л.В. Лютикова

Abstract: Methods for solution of a large class of problems on the base of multi-valued and probabilistic logics have been discussed. Logic and network approach for analogous derivation has been suggested. Method for regularity search, logic-axiomatic and logic-probabilistic methods for learning of terms and pattern recognition in the case of multi-valued logic have been described and generalized. Defeasible analogical inference applications and new forms of inference using exclusions are considered. The methods are applicable in a broad range of intelligent systems.

Введение

Классическая двузначная логика связана с формализацией строго корректных (формальных) рассуждений. Однако очень часто предметная область, на базе которой строятся основные понятия и выводы, обладает неполной, неточной, противоречивой и зачастую изменчивой информацией [1-7]. В связи с этим возникает необходимость в использовании и развитии новых неклассических методов формализации интеллектуальных процессов и информационных технологий.

На сегодняшний день арсенал разнообразных неклассических логик достаточно велик [2,3,7]. Однако плохо разработана методика использования этих логик в конкретных задачах. Кроме того, возможности этих логик (например, К-значных логик) не полностью отвечают потребностям, возникающим при разработке интеллектуальных систем и технологий.

До сих пор остаётся популярным статистический подход к анализу данных и принятию оптимальных решений. Однако он требует репрезентативности исходных данных. На практике обучающие выборки данных, из которых извлекаются знания и формулируются интеллектуальные решения, очень ограничены и поэтому не являются статистически репрезентативными.

В докладе рассматриваются способы применения многозначных и вероятностных логик к решению интеллектуальных задач (в частности, к задачам обучения и поиска закономерностей на примере трёхзначных логик). Используется несколько подходов к построению вывода на основе модифицируемой информации: вывод по аналогии, логико-аксиоматический и логико-вероятностный методы и моделирование сетевых потоков. Показано, что даже в такой разработанной области как вывод с использованием исключений применение неклассических логических средств позволяет значительно расширить как область приложения, так и теоретический базис разработки.

Предложенные методы позволяют кооперировать логические и вероятностные подходы и получать преимущества от каждого из них.

1. Определение классов и вывод по аналогии

Пусть универсум классов V составляют подмножества S_1, S_2, ... и $S \in V$. Каждое из подмножеств типа S включает элементы $x_{s:1}$;, $x_{s;2}$, ..., которые формируют новую модель. Исходное множество S относится к одному из классов $S_i \in V$. Конечный результат анализа S идентифицируется с одним из классов S_i в U. На выходе формируется ответ типа V_s = (T; F;?) со тремя значениями: “истина”, “ложь”, “неопределенность”. В случае ответа V_s =? или V_s=F множество S может быть идентифицировано с больше, чем одним из известных классов S_{i1}, S_{i2}, ... ($i1 \neq i2$ ...). Ответ V_s=T получается тогда и только тогда, когда изучаемый класс S совпадает с S_i.

* Работа выполнена при частичной поддержке грантов РФФИ №03-01-00224а и гранта РГНФ №03-06-12019в

Среди классов S_i существует зависимость типа " предшественник - наследник ", (например S_i - предшественник S_{i1}). Таким образом, можно формировать семантические сети. Необходимо заметить что элементы $x_{si1;1}$, $x_{si1;2,}$ ... дают различия между каждым классом S_{i1} и другими наследниками общего наследника S_i. Все различия, которые выступают в процессе сравнения S_{i1} с другими классами, не являющимися прямыми наследниками S_i, определяются после использования механизма наследственности.

Вывод (ответ) V_s=T формируется, когда для всех предшественников S_i и для S_{i1}, соответствующие конъюнктивные члены имеют следующий вид: $A_1' \Lambda A_2' \Lambda .. A_n'$, где A_k есть x_k или $\neg$ x_k; A_k' может совпадать с A_k или включать (с использованием дизъюнкции) A_k и аналогичные члены для других переменных.

Пусть импликативные правила типа Хорна описывают некоторую предметную область:

$$B \leftarrow \underset{i \in I}{\Lambda} A_i . \qquad (1)$$

При использовании двузначной логики в указанных правилах, если по крайней мере одна переменная A_i - не “истина”, то истинность B не определена, т.е. B может иметь значения “истина” или “ложь”. В случае, если соответствующее исключение из конъюнкции (1) правила - основываются на подключении члена с любым A_k ($k \in I$), процедура для вывода изменяется. В случае, если исключение E (C, A_k) и C является истинным и A_k ложным, правая часть правила B может быть истинна (как исключение).

Расширенные модели вывода с исключением были предложены, обобщены в формализованных в [9,10] в следующем виде.

$$\frac{B \leftarrow \overset{z}{\underset{i=1}{\Lambda}} A_i C, E(C, A_k), \neg A_k \leftarrow C}{B \leftarrow A_1 \Lambda A_2 \Lambda \ldots A_{p-1} \Lambda \neg A_k \Lambda \ldots A_z} , \qquad (2)$$

$$\frac{C,\ B \leftarrow \overset{z}{\underset{i=1}{\Lambda}} \mathrm{A}_i ,E(C, A_k)}{B \leftarrow A_1 \Lambda \ldots A_{k-1} \Lambda A_{k+1} \Lambda \ldots A_я} , \qquad (3)$$

$$\frac{C,\ B \leftarrow \overset{z}{\underset{i=1}{\Lambda}} \mathrm{A}_i ,\ \ E(C, A_k)}{B \leftarrow A_1 \Lambda \ldots A_{k-1} \Lambda A_{k+1} \Lambda (A_k V C) \Lambda A_k \ldots A_z} \qquad (4)$$

Из формул (2)-(3) видно, что исключения - это своего рода подключение специальных правил, и области их действия. Интерпретация формулы (2) основана на следующем: если существует исключение E(C, A_k), связанное одним из правил с заключением B и в результате его воздействия получен A_k, то конъюнкт A_k должен быть заменен на $\neg A_k$. В случае, если C - не “истина”, то соответствующая замена не выполнима. При обращении к правилу Modus Ponens, связь между B и $\neg A_k$ ведет к формальному логическому противоречию.

Следовательно, формирование исключений типа E(C, A_k) может приводить к противоречивому результату, вызванному неполнотой в описании предметной области. В случае, если C истинно, исключение E (C, A_k) подключает это значение к конъюнктору A_k, чтобы обнулить значение последних выводов. В результате A_k заменяется на C, так как проверка его значения не влияет на вывод. В случае, когда C истинно, соответствующий конъюнкт A_k прямо заменяется на C.

Правила вида (1) объединяются в системы:

$$\begin{cases} B_1 \leftarrow \underset{i \in I}{\Lambda} A_{1i} . \\ B_2 \leftarrow \underset{j \in I}{\Lambda} A_{2j} . \\ \ldots \end{cases} \qquad (1A)$$

В общем случае причинно-следственная связь может быть реализована с использованием неклассических операции следования, обозначенных в докладе ‘<-‘, а B_i могут быть представлены в виде комбинации сложных логических связей:

$$\begin{cases} B_1 <- \underset{i \in I}{\Lambda} A_{1i}. \\ B_2 <- \underset{j \in I}{\Lambda} A_{2j}. \\ \ldots \end{cases} \qquad (1Б)$$

Использование исключений (2)-(4) может применяться и в системах (1А) или (1Б) и в общем случае представляет изменение соотношений между различными частями причинно-следственных связей под действием новой информации (исключение, прикрепленное к одной или группе связей). Новая информация может изменить взаимосвязь между элементами правила (1) или системы (1А) ; (1Б). В этом случае связи причинно-следственного характера разрываются или усиливаются под действием дополнительной информации, содержащейся в исключениях. Далее в докладе не будут обсуждаться варианты (1А) и (1Б), так как в большинстве наших практических приложений достаточно ограничиться правилами (1), при этом алгоритмическая сложность используемой комбинации методов падает в значительной мере. По своей природе представленные исключения представляют расширенный вариант defeasible inference, широко используемый в интеллектуальных системах. В отличие от классического вывода с использованием исключений, в представленной работе возможно не только удаление содержащегося в правиле и прикрепленном к нему исключении A_k, но и включение в правило новой подформулы – например $\neg A_k$ в формуле (2) или взаимосвязи между A_k и C в (4). Также исследованы и варианты формул с использованием неклассического отрицания ~, вариации с устранением импликации под действием исключения и другие:

$$\frac{B \leftarrow \overset{z}{\underset{i=1}{\Lambda}} A_i C, E(C, A_k), \sim A_k \leftarrow C}{B \leftarrow A1 \wedge A2 \wedge \ldots A_{p-1} \wedge \sim Ak \wedge Ak1 \ldots Az}, \qquad (2А)$$

$$\frac{C,\ B \leftarrow \overset{z}{\underset{i=1}{\Lambda}}\ A_i\ , E(C, Ak)}{A1 \wedge \ldots A_{k-1} \Lambda A_{k+1} \Lambda \ldots Aя}, \qquad (3А)$$

$$\frac{C,\ B \leftarrow \overset{z}{\underset{i=1}{\Lambda}}\ A_i\ , E(C, Ak)}{A1 \wedge \ldots A_{k-1} \Lambda A_k \Lambda A_{k+1} \Lambda \ldots Aя}, \qquad (3Б)$$

где A_{k1} – дополнительное условие для перехода от $\sim A_k$ к $\neg A_k$. Также исследованы схемы многоаргументных исключений $E(C,A_k,A_l,\ldots A_s)$, приводящие к одновременному изменению нескольких частей правила. Предложенный метод приводит к трем основным результатам: к изменению истинности частей правила под действием исключения (если исполнены условия для активизации исключения), к введению или исключению подформул из соответвующего правила или к разрушению самого правила, как показано в (3А) или (3Б). Результаты исследований привели к большому количеству вариаций вывода с исключениями, часть из которых представлена в цитируемых наших работах.

Для увеличения эффективности поиска предлагается использовать вывод по аналогии. Этот метод подробно изложен в [8-10].

2. Сетевые потоки и вывод по аналогии

В интеллектуальных системах важную роль играют графовые модели и сетевые потоки. Пусть задан граф G(N, U) с множеством ребер U и множеством вершин N. В [10] показано, что вывод по аналогии может быть представлен как сетевой поток на графе. Геометрическая интерпретация этого представления изображена на рис.1

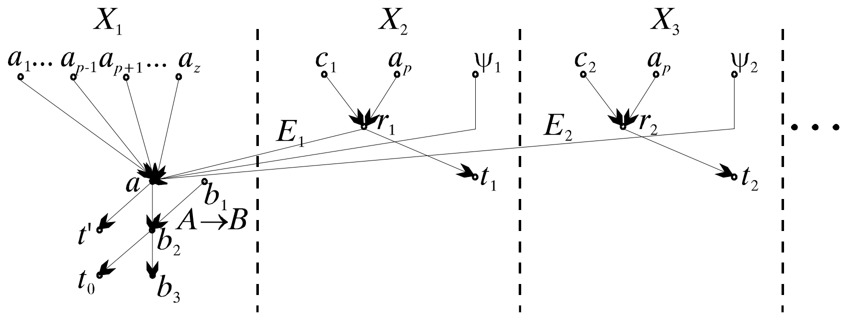


Рис. 1.

Рис.1 разделён пунктирами на области. Каждая область содержит данные, соответствующие приблизительно одному объектуXj. Множество S содержит все элементы A_i, исключения E и ψ_i. Множество заключений T содержит все t_ν, t_0, и t'. Тогда для всех X,y∈N справедливы следующие соотношения:

$$f(y,X)-f(X,y)=\begin{cases} \nu(a_j), \text{ если } y\in S, \\ 0, \text{ если } y\notin S,T, \\ \nu(t_k), \text{если } y\in T. \end{cases} \qquad (5)$$

Здесь A_i соответствует потоковой функции $f(a_i,a)$, а C_ν, A_p и $E(C_\nu,A_p)$ - функциям $f(c_i,r_i)$, $f(a_p,r_i)$, $f(r_i,a)$ соответственно. Функция $f(\psi_i,a)$ первоначально принимает значение 1, если $\psi(X_j,X_1)<T$ или значение 0 - в противном случае. Некоторые соответствующие исключения $E(C_\nu,A_p)$, i≥j могут быть включены в знания об объекте X_j. Пары входящих ребер дизъюнктивно соединены в вершинах r_i, и конъюнктивно соединены в вершинах *a* и b_2. Функциональная зависимость v выглядит следующим образом:

$$\nu(a_j)=f(a_j,a),\ \nu(tj)=f(a,t_j). \qquad (6)$$

Конъюнкция всех A_i, и ψ_i обозначается через A и соответствуют ребру (a,b_2). Импликация A→B есть множество ребер (b_1,b_2), а результат дедуктивного вывода $f(b_2,b_3)$ имеет значение истинности B. Тогда вывод по аналогии можно представить следующей системой равенств и неравенств [6-10]:

$$f(a,b_2)-f(a_ia)\le 0;\ i=1,...z, \qquad (7)$$

$$f(a,b_2)-f(r_j,a)\le 0;\ j=1,...n, \qquad (8)$$

$$f(a,b_2)-f(r_j,a)\le 0;\ j=1,...n, \qquad (9)$$

$$f(r_j,a)-f(c_j,r_j)\ge 0;\ j=1,...n, \qquad (10)$$

$$2f(r_j,a)-f(c_j,r_j)-f(a_p,r_{1j})=0;\ j=1,...n, \qquad (11)$$

$$(2n+z-1)f(a,b_2)-\sum_{i=1}^{z-1}f(a_i,a)-\sum_{j=1}^{n}f(r_j,a)-\sum_{j=1}^{n}f(\psi_j,a)=0, \qquad (12)$$

$$2f(b_2,b_3)-f(a,b_2)-f(b_1,b_2)=0, \qquad (13)$$

$$f(r_j,a) = 0 \text{ или } 1, \qquad (14)$$

$$f(x,y)\ge 0;\quad (x,y)\in U, \qquad (15)$$

$$f(r_j,t_j)\le 1, \qquad (16)$$

$$f(a,t')\le 2n+z-2, \qquad (17)$$

$$f(b_2,t_0)\le 1.$$

Таким образом, проблема доказательства по аналогии сведена к задаче линейного программирования с целевой функцией вида:

$$\sum_{(x,y)\in D} f(x,y) \to max \qquad (18)$$

при ограничениях (7) – (17). Более подробно эта задача описана в [10]

3. Логический вывод в задачах поиска закономерностей в базах данных и распознавания образов

Предположим, что информация о некоторой предметной области задана в виде базы данных, интерпретируемой как обучающая выборка для поиска (извлечения) логических закономерностей, связывающих эти данные. Пусть множество $Z=\{X_i,Y_i\}^m_{i=1}$ является некоторой базой данных (обучающей выборкой), связанных неизвестной зависимостью вида

$$Y = f(X), \qquad (19)$$

где X и Y являются многозначными предикатами. Требуется определить зависимость (закономерность) (19) по обучающей базе данных Z мощности m.

Сначала рассмотрим случай кодирования обучающей выборки двузначными предикатами. В этом случае исходная предметная область может быть описана правилами продукций вида:

$$\Lambda^n_{j=1} x_{ij} \to y_i, j = 1,..., m \ . \qquad (20)$$

Поскольку любое правило продукции является импликацией, оно может быть представлено в виде совершенной дизъюнктивной нормальной формы (СДНФ). В случае двузначной логики правило перевода импликации в ДНФ осуществляется по формулам:

$$A \to B = \neg A V B \to \qquad (21)$$

Поэтому в случае кодирования знаний двузначными предикатами любое предложение может быть представлено правилом продукции и переведено в СДНФ вида:

$$V^n_{i=1} x_i^{\sigma_{ij}} V y_i \ , \qquad (22)$$

где σ_{ij} принимает значения 0 или 1.

Далее требуется объединить все формулы обучающей выборки в одну логическую функцию или систему функций, дающих однозначную интерпретацию исходной предметной области. Таким образом, неизвестная зависимость Y=f(X) может быть восстановлена непосредственно по обучающей выборке Z.

Любая логическая функция, записанная в виде СДНФ, может быть подвергнута сокращению. Поэтому и система логических знаний, как правило, тоже может быть подвергнута сокращению. Тогда сокращение СДНФ, соответствующей логической функции, может интерпретироваться как минимизации исходной базы знаний.

Для сокращения СДНФ с учётом специфики предметной области предлагается следующий алгоритм:

1. Если в ДНФ имеются однолитерные дизъюнкты x и $\neg x$, то ДНФ общезначима;

2.Если некоторая переменная входит в ДНФ с одним знаком, то удаляем все дизъюнкты, содержащие эту переменную (данная переменная неинформативна);

3.Если в ДНФ имеется какой-то однолитерный дизъюнкт x, то выполняем следующие действия:

а) удаляем все дизъюнкты вида $x \Lambda$ ... (правило поглощения);

б) заменяем дизъюнкты вида $\neg x \Lambda s$ ... на дизъюнкты вида $s \Lambda p$ ... .

В результате такого сокращения получаем самые “сильные” логические правила, описывающие исходную предметную область.

Описанный метод можно использовать для обучения понятием (классом) в задачах распознавания образов. Синтезированные понятия можно интерпретировать как аксиомы классов (образов) $A_k(\omega)$ в предметной области, заданной обучающей базой данных. Тогда задача распознавания образов сводится к поиску логического вывода с использованием метода резолюций Робинсона или обратного метода Маслова [11].

Задача идентификации изображения $\hat{\omega}$ k-го класса (образа) на сложном изображении ω с логическим описанием D(ω) сводится к выводу формулы:

$$D(\omega) \to \exists\hat{\omega} A_k(\omega)\,,\ \ \hat{\omega} \in \omega \tag{23}$$

Смысл этой формулы заключается в следующем: сложное изображение ω, имеющая логическое описание D(ω), содержит изображение $\hat{\omega}$ k-го класса, на котором истинна аксиома $A_k(\hat{\omega})$. Это позволяет автоматически идентифицировать и локализовать (выделить) изображение k-го класса (образ) на сложном изображении, содержащем изображения (образы) из M разных классов $S_1, S_2, \ldots, S_M$.

Многократное применение логико-аксиоматического метода для каждого k=1,2,…,M позволяет распознать (классифицировать) все изображения всех классов, содержащиеся на сложном изображении [11].

4. Использование многозначных и вероятностных логик в задачах обучения и поиска закономерностей

Описанный метод поиска логических закономерностей может быть обобщён на случай многозначного кодирования обучающей выборки и поиска многозначных закономерностей. Использование многозначных логик осложняется неоднозначностью интепретации функций отрицания, импликации и т.д. Поэтому рассмотрим наиболее общий вариант в случае использования трёхзначной логики.

Пусть множество значений истинности имеет вид{0 1 2} со следующей интерпретацией:

x=0 –ложь, *x*=1-бессмыслица (неопределённость), *x*=2-истина.

Тогда введем понятие инверсии как $\neg x$= 1V0, т.е отрицание истинны может быть либо ложью, либо бессмыслицей. Это понятие определено таблицей 1. Такое задание инверсии обеспечивает включение всех возможных интерпретаций инверсии в разных логиках.

Таблица 1

X	¬X
0	1V2
1	0V2
2	0V1

Введем еще несколько функций трехзначной логики. Наиболее важными из них являются характеристические функции, определяемые следующим образом:

$$I_i(x) = \begin{cases} k-1, & если\ x = i, \\ 0, & если\ x \neq i, \end{cases} \tag{24}$$

$$J_i(x) = \begin{cases} 1, & при\ x = i, \\ 0, & при\ x \neq i, \end{cases} \tag{25}$$

Основные правила операций с этими функциями имеют вид:

$$I_\sigma(x) I_\tau(x) = \begin{cases} I_\sigma(x), & если\ \sigma = \tau, \\ 0, & если\ \sigma \neq \tau, \end{cases} \tag{26}$$

$$\sigma \wedge \tau = \min(\sigma,\tau), \qquad \sigma \vee \tau = \max(\sigma,\tau). \tag{27}$$

Воспользуемся также двузначным аналогом импликации в данной трехзначной логике, т.е.

$$A \rightarrow B = \neg A \vee B = I_0(A) \vee I_1(A) \vee B \quad (28)$$

Форма (28) так же, как инверсия, является расширением, включающим в себя ряд возможных импликаций трехзначной логики. Такое широкое задание основных функций логики удобно при моделировании интеллектуальных систем в тех случаях, когда не представляется возможным описать интеллектуальные процессы какой-либо одной конкретной многозначной логикой.

Теперь вернемся к решению исходной задачи в терминах трехзначной логики. Каждую строку обучающей выборки опишем правилами продукции:

$$\Lambda^{n}_{j=1} X_{ij} \rightarrow Y_i, \; i = 1,..., \; m \;. \quad (29)$$

Тогда аналогом СДНФ будет следующая функция трехзначной логики:

$$I_0(I_l(x)) = \begin{cases} I_1(x).. \vee I_{l-1}(x) \vee I_{l+1}(x).. \vee I_m(x), если \quad x = l, \\ 0 \; если \; x \neq l \end{cases} . \quad (30)$$

$$V^{m}_{j=1} I_0(I_i(x)) \vee Y_i, \quad (31)$$

Поскольку каждую закономерность (знание), согласующееся с обучающей выборкой, можно записать в виде предложенной функции трехзначной логики, хотелось бы иметь возможность представления всех закономерностей, образующих базу знаний, функцией или системой функций трехзначной логики.

Однозначное соответствие легко получить, если, например, логически перемножить правила продукций. Это соответствует рассуждениям следующего типа: знаем частные (локальные) правила и таким образом знаем все вместе локальные правила (закономерности), определяющие глобальную базу знаний, построенную по обучающей выборке.

В результате получим трёхзначную функцию, определяющую искомую закономерность. К этой функции можно применить алгоритм сокращения в адаптированном для многозначных логик варианте, а именно:

1. Если некоторая переменная входит в ДНФ с одним знаком ($I_j(x)$, j=const, во всех дизъюнктах), то удаляем все дизъюнкты, содержащие эту переменную (данная переменная неинформативна);

2. Если в ДНФ имеется какой-то одно-литерный дизъюнкт $I_j(x)$, то выполняем следующие действия:

   а) удаляем все дизъюнкты вида $I_j(x)\ \Lambda$ ... (правило поглощения);

   б) заменяем дизъюнкты вида $I_i(x)\ \Lambda$ s... ($i \neq j$) дизъюнктами вида s Λ p ... .

Результатом использования алгоритма является многозначная функция, построенная по исходной обучающей выборке, однозначно её характеризующая и дающая множество наиболее существенных правил (закономерностей), определяющих исходную область знаний.

При добавлении нового правила продукции (нового знания) проверяем, выводимо ли данное правило из уже существующих или нет. Если это правило выводимо, то функция остается той же. В противном случае к базе знаний присоединяем новое правило (закономерность) путем логического многозначного перемножения имеющейся функции и нового правила продукции, записанного в виде многозначной СДНФ.

Другой метод обучения понятиям и поиска многозначных закономерностей по заданным базам данных основан на локально-оптимальных логико-вероятностных алгоритмах [12,13]. Этот метод обеспечивает автоматический синтез, оптимизацию (по точности) и минимизацию сложности баз знаний в терминах многозначных предикатов с произвольной значностью по обучающим базам данных. Он допускает интерпретацию и реализацию синтезированных знаний (закономерностей) в виде трёхслойных или многослойных нейронных сетей полиномиального типа с самоорганизующейся архитектурой [14,15].

Заключение

Предложен подход для вывода по аналогии на базе трёхзначной логики и сетевых потоков, ориентированный подход для применения в системах искусственного интеллекта и поддержки принятия решений. Обсуждены особенности и общие характеристики различных типов доказательства по аналогии.

Разработан метод, в котором отменяемое доказательство выполнено в форме сетевого потока. При этом подходе проблемы логического программирования сводятся к соответствующим проблемам линейного программирования.

Предложен многозначный логический подход к решению задач обучения и поиска закономерностей в базах данных, позволяющий однозначно описывать предметную область, структурно анализировать исходную информацию, сокращать её и изменять по мере формирования новых знаний, не выводимых из исходных данных.

Логико-аксиоматический и логико-вероятностный методы обучения понятиям и распознавания образов обобщены на случай многозначных логик. Показано, что синтезируемые понятия и распознающие правила могут быть реализованы в виде многозначных нейронных сетей полиномиального типа и использованы в системах интеллектуального и нейронного управления [13-14].

Литература

[1] P.Dung, P. Moncarella, Production systems with negation as failure, IEEE, Trans/ Knowledge and Data Engineering, vol. 14, no.2, pp. 336-352, April 2002.

[2] N. Rescher, Many Valued Logics, Basil Blackwell, Oxford 1979.

[3] Яблонский С.В. Введение в дискретную математику. Москва. Высшая Школа, 2001

[4] V. Gladun, Process of new knowledge formation. Pedagog 6, Sofia, 1994.

[5] Z. Markov, Inductive Machine Learning Methods, TEMPUS JEN 1497, Softex, Sofia, 1996.

[6] V. Jotsov, Knowledge discovery and data mining in number theory: some models and proofs, Proc. Methods and Algorithms for Distributed Information Systems Design. Institute for Information Transmission Problems of RAS, 1997, pp.197-218.

[7] K. Kowalski, Logic for Problem Solving, North-Holland Inc., Amsterdam, 1979.

[8] Halpern, J., M. Rabin, A logic to reason about likelihood, Artificial Intelligence, 1987, no. 32, pp. 379-405.

[9] V.Sgurev, V.Jotsov, Some characteristic features of the application of three-valued logical schemes in intelligent systems, Proc. First International IEEE Symp. 'Intelligent Systems' (T. Samad and V. Sgurev, Eds.), Vol. I., Varna, Bulgaria, September 10-12, 2002, pp. 171-176.

[10] V.Sgurev, V.Jotsov, Some defeasible inference methods by using network flows, J. Problems of Engineering Cybernetics and Robotics, 2001, no. 51, pp. 110-116.

[11] A.Timofeev, T.Kossovskaya, Logic-Axiomatical Method for Knowledge Representation and Recognition in Intelligent Manufacturing Systems, Elsevier, Amsterdam – Oxford – New York, 1990, pp. 3-6.

[12] A.Timofeev, Z.Shibzuchov, The Synthesis and Optimization of Knowledge Bases Based on the Local-Optimization Logic-Probabilistic Algorithms. Int. Journal of Information Theories and Applications, 1995, vol. 3, no. 2.

[13] A.Timofeev, R.Yusupov, Evolution of Intelligent Control in Adaptive Systems. Int. Journal of Adaptive Control and Signal Processing, 1992, vol. 6, pp.193-200.

[14] Тимофеев А.В., Шеожев А.В., Шибзухов З.М. Синтез нейросетевых архитектур по многозначному дереву решений. – Нейрокомпьютеры: разработка и применение, 2002, № 5-6, с. 44-49.

[15] Timofeev A.V. Polynomial Neural Networks with Self-Organizing Architecture, Int. Journal on Optical Memory and Neural Networks, 2004, v.1.

Информация об авторах

В.С. Сгурев, В.С. Йоцов: Институт информационных технологий Болгарской академии наук;
P.O.Box 161, Sofia 1113, BULGARIA; sgurev@bas.bg, jotsov@ieee.org

А.В. Тимофеев, Л.В. Лютикова: Санкт-Петербургский институт информатики и автоматизации РАН,
199178 -Санкт-Петербург, 14-я линия, д.39.; tav@iias.spb.su

АДАПТИВНОЕ УПРАВЛЕНИЕ И САМООРГАНИЗАЦИЯ В МУЛЬТИ-АГЕНТНЫХ ИНФОКОММУНИКАЦИОННЫХ СЕТЯХ

А.В.Тимофеев

Введение

Важную роль для поддержки и развития "всемирного диалога" между людьми будут играть не только глобальная сеть Internet, но и более совершенные компьютерные и телекоммуникационные сети (ТКС) нового поколения, создаваемые в проектах "Internet 2", "Abilene", "NGI" и т.п.

Проектирование ТКС нового поколения требует разработки принципиально новых подходов к сетевому управлению и самоорганизации на основе теории адаптивных и интеллектуальных систем управления и мульти-агентных и нейросетевых технологий.

В докладе обсуждают общие проблемы, возникающие в глобальных ТКС нового поколения, и предлагаются некоторые новые методы мульти-агентного управления и сетевой самоорганизации.

1. Архитектура мульти-агентных инфокоммуникационных сетей

Глобальные мульти-агентные ТКС служат для представления пользователям информационных и вычислительных ресурсов, распределённых в компьютерных сетях (КС).

Архитектура таких ТКС состоит из четырёх основных (базисных) подсистем:

1. Распределённая система связи (РСС);
2. Сетевая система управления (ССУ);
3. Распределённая информационная система (РИС);
4. Распределённая транспортная система (РТС).

Эти подсистемы связаны между собой и предназначены для управляемой передачи пользователям глобальной ТКС распределённых информационных и вычислительных ресурсов, хранящихся в глобальной КС, охватывающей многие страны.

РСС состоит из распределённых средств доступа и пользовательских интерфейсов, а также портов и шин данных, обеспечивающих прямую и обратную связь между пользователями глобальной ТКС, различными подсистемами ТКС и связанной с ней глобальной КС, состоящей из распределённых (удалённых на значительные расстояния) компьютеров или локальных КС.

В качестве пользователей глобальной ТКС могут выступать клиенты, администраторы, операторы и провайдеры ТКС.

ССУ получает через РСС запросы клиентов и команды сетевых администраторов ТКС, обрабатывает внутреннюю информацию о текущем состоянии РТС и внешнюю информацию о состоянии информационных и вычислительных ресурсов в КС, поступающую от РИС, и формирует управление РТС, обеспечивающее удовлетворение запросов пользователей путём передачи им необходимых информационных и вычислительных ресурсов КС.

РИС получает сигналы внутренней и внешней обратной связи о текущем состоянии РТС как объекта управления и доступных информационных и вычислительных ресурсов, хранящихся в глобальной КС. Она передаёт эти сигналы в ССУ для формирования или коррекции управления потоками данных, передаваемых через РТС.

РТС состоит из коммуникационных узлов (в роли которых могут выступать специальные коммуникационные процессоры) и каналов связи между ними. Она играет роль распределённого объекта управления и служит для управляемой адресной передачи потоков данных от пользователей к КС через ТКС и обратно.

Все указанные подсистемы глобальной ТКС имеют распределённый характер, взаимосвязаны и активно взаимодействуют между собой в процессе обслуживания пользователей информационными и вычислительными ресурсами, хранящимися в глобальной КС.

2. Самоорганизация и адаптация в сетевом управлении

Главную роль в целенаправленной самоорганизации и высококачественной обработке информации и адресной передаче потоков данных по запросам пользователей играет ССУ.

Основной задачей ССУ глобальных ТКС нового поколения, работающих на больших скоростях передачи данных, является самоорганизация и автоматическое формирование такого управления потоками данных в РТС, которое поддерживает трафик гетерогенных данных большого объёма с надёжными гарантиями высокого качества обслуживания (Quality of Service, QoS) пользователей ТКС.

Решение этой глобальной задачи сетевого управления и самоорганизации в ТКС распадается на локальные задачи управления потоками данных, адаптации к изменяющемуся трафику, предотвращения перегрузок, разрешения сетевых конфликтов и т.п. Практическая реализация решения этих задач осуществляется с помощью специальных сетевых протоколов передачи информационных и управляющих сигналов и потоков данных.

В общем случае указанные проблемы сетевого управления и самоорганизации должны решаться для двух основных платформ глобальных ТКС:

- - объединённые IP-сети, взаимодействующие через маршрутизаторы потоков данных по протоколу IP (Internet Protocol) из набора протоколов TCP/IP (Transmission Control Protocol/Internet Protocol);
- - ATM-сети, использующие протоколы ATM (Asynchronous Transfer Mode).

Сегодня эти платформы активно развиваются и конкурируют на рынке сетевых инфокоммуникационных услуг, что проявляется в так называемой “битве между IP и ATM”. В этой связи представляется особенно важной такая эволюция ССУ, которая обеспечит конвергенцию и интеграцию IP- и ATM-сетей в глобальных ТКС нового поколения и их дальнейшее развитие.

Традиционно для организации управления потоками данных и оборудованием РТС используются сетевые принципы и архитектуры централизованного или децентрализованного управления. Каждый из этих принципов обладает определёнными достоинствами и недостатками.

Централизованная архитектура ССУ основывается на выделении центрального компьютера, связанного через РСС с администратором ТКС и выполняющего функции “центра управления” передачей данных через узлы и каналы связи РТС.

Достоинством такой архитектуры является глобальность управления из единого “центра”. Недостатками централизованного управления являются отсутствие самоорганизации и низкая надёжность и отказоустойчивость, проявляющиеся в том, что выход из строя центрального управляющего компьютера приводит к полной или частичной потере управляемости РТС. Поэтому обычно предусматривается резервирование центрального компьютера ССУ.

Децентрализованная архитектура ССУ распределяет функции обработки информации и управления между рядом локальных компьютеров, управляющих различными сегментами РТС или потоками данных в них.

Достоинством такой архитектуры является то, что относительная независимость распределённых локальных “центров управления” повышает надёжность адресной передачи потоков данных. Недостатками децентрализованного управления являются локальность и неполнота целей управления, что требует координации и согласованной работы распреде-лённых локальных управляющих компьютеров.

3. Сетевые агенты и мульти-агентное управление

Учитывая недостатки описанных традиционных сетевых архитектур, целесообразно разработать “гибридную” самоорганизующуюся архитектуру ССУ для глобальных ТКС нового поколения, сочетающую в себе достоинства централизованной и децентрализованной архитектур. Будем называть такую "компромиссную" самоорганизующуюся архитектуру мульти-агентной архитектурой ССУ.

В этом случае основные функции обработки информации, самоорганизации и управления потоками данных в глобальных ТКС нового поколения распределяются между взаимосвязанными интеллектуальными или нейросетевыми агентами.

Каждый сетевой агент имеет собственную локальную БД и БЗ или нейронную сеть (НС) с самоорганизующейся архитектурой и средства связи с другими агентами для обмена информацией в процессе совместного (кооперативного) принятия решений, самоорганизации и автоматического формирования сетевого управления РТС, обеспечивающего адресную доставку информационных вычислительных ресурсов КС по запросам пользователей глобальной ТКС.

В роли сетевых агентов могут выступать компьютеры или нейросетевые маршрутизаторы ССУ, связанные с узлами РТС, а также программные агенты РСС и РИС. Будем называть таких агентов внутренними агентами глобальной ТКС нового поколения. Тогда роль внешних агентов играют пользователи (клиенты, администраторы и т.п.) вместе со средствами доступа в ТКС и сетевым интерфейсом, а также компьютерные узлы (хосты) распределённой КС.

В процессе проектирования ССУ на базе теории агентов и принципов самоорганизации возникают новые проблемы многоадресной и многопотоковой маршрутизации и мульти-агентного диалога между внутренними агентами глобальной ТКС нового поколения, внешними агентами-пользователями и агентами КС как распределённого мирового хранилища данных, знаний и приложений.

Решение этих проблем требует разработки методов самоорганизации, включающих в себя средства предотвращения или автоматического разрешения сетевых конфликтов под контролем адаптивной или интеллектуальной ССУ с мульти-агентной архитектурой .

Для управляемой адресной передачи и навигации потоков данных, разрешения сетевых конфликтов, функциональной диагностики и распознавания состояний глобальной ТКС нового поколения целесообразно ввести специальных агентов-координаторов (например, на уровне маршрутизации потоков данных) и, возможно, других глобальных агентов, обеспечивающих самоорганизацию.

Особенность этих агентов высокого уровня заключается в том, что их БД и БЗ формируются на основе локальных БД и БЗ агентов более низкого уровня. Поэтому они имеют глобальный (мульти-агентный) характер и позволяют оценивать сетевую ситуацию и осуществлять самоорганизацию "в целом".

Таким образом, развитие и совершенствование самоорганизующихся архитектур ССУ глобальных ТКС нового поколения должно осуществляется не только и не столько "вширь" (т.е. "по горизонтали" охвата территории), но и "вглубь" (т.е. "по вертикали" эволюции иерархии сетевого управления и самоорганизации).

Важную роль при этом играют процессы адаптации, самоорганизации и интеллектуализации ССУ.

4. Проблемы адаптивной маршрутизации информационных потоков

Рассмотрим основные особенности процессов адаптации и самоорганизации на примере адаптивной мульти-агентной маршрутизации потоков информации в ТКС.

Необходимость в адаптивной маршрутизации возникает при непредсказуемых изменениях структуры (узлов и каналов связи) ТКС или при перегрузке буферов узлов или каналов связи ТКС. По существу речь идет о маршрутизации и самоорганизации потоков информации в нестационарных глобальных ТКС с переменной структурой и нагрузкой.

Причинами изменения структуры ТКС могут быть как добавление или отказ (выход из строя) отдельных узлов и каналов связи, так и сетевые перегрузки, препятствующие передаче потоков данных по запрещенным (перегруженным) узлам и каналам. Поэтому маршрутизаторы должны планировать и корректировать оптимальные маршруты передачи пакетов данных, адаптируя их к возможным изменениям ТКС, происходящих в реальном времени. Для этого необходима обратная связь о текущем состоянии узлов и каналов связи ТКС, которая может быть организована путем обмена информацией между узлами ТКС.

Отличительными чертами адаптивной маршрутизации по сравнению со статической или динамической маршрутизацией являются следующие особенности:

- алгоритмы адаптивной маршрутизации требуют учета и обработки текущей информации о реальном состоянии ТКС, что делает их более сложными и увеличивает время определения оптимального маршрута;
- передача информации о состоянии или структурных изменениях в ТКС к адаптивным маршрутизаторам дополнительно загружает сеть и приводит к задержкам (запаздыванию);
- увеличение сетевой нагрузки и времени запаздывания может приводить к колебаниям или автоколебаниям и увеличению числа шагов при определении оптимального маршрута.

Адаптивная маршрутизация потоков данных в глобальных ТКС имеет ряд преимуществ по отношению к неадаптивной (статической или динамической) маршрутизации, а именно:

- обеспечивает работоспособность и надежность ТКС при непредсказуемых изменениях их структуры или параметров;
- приводит к более равномерной загрузке узлов и каналов связи ТКС за счет «выравнивания» нагрузки;
- упрощает управление передачей потоков данных и облегчает адаптацию к сетевым перегрузкам;
- увеличивает время безотказной работы и производительность ТКС при высоком уровне предоставляемых услуг в непредсказуемых условиях изменения сетевых параметров и структуры, что особенно важно для агентов-пользователей ТКС.

Достижение этих преимуществ в значительной степени зависит от используемых принципов и алгоритмов адаптивной маршрутизации и самоорганизации потоков данных в ТКС с непредсказуемо изменяющейся структурой и заранее неизвестным трафиком.

Как отмечается в 6-ом издании монографии Tanenbaum A.S. Computer Networks (Prentice Hall, 1996), «адаптивная маршрутизация – это задача, которую весьма трудно решить должным образом. Доказательством этого может служить тот факт, что наиболее крупные сети с пакетной коммутацией (такие, как ARPANET и ее «наследники», TYMNET и сетевые архитектуры IBM и DEC) неоднократно претерпели значительные изменения принципов маршрутизации».

5. Методы адаптивной и мульти-агентной маршрутизации

Принципы адаптивной маршрутизации и самоорганизации потоков данных можно разбить на три класса в зависимости от используемой информации о реальной (текущем) состоянии ТКС, т.е. от характера сигналов обратной связи:

- локальная информация (обратная связь) от одного узла ТКС;
- локальная информация (обратная связь) от узла и его «соседей» в ТКС;
- глобальная информация (обратная связь) от всех узлов ТКС;

Простейший принцип адаптивной маршрутизации с локальной обратной связью от одного узла заключается в том, что пакет данных передается в канал связи с самой короткой очередью или с наибольшей вероятностью предпочтительности канала. При этом может происходить локальное выравнивание нагрузки в выходных каналах ТКС. Однако в этом случае возможно отклонение от оптимального маршрута.

Более эффективные принципы адаптивной маршрутизации основываются на передаче в начальный узел локальной информации (обратной связи) от соседних узлов или глобальной информации от узлов ТКС. В качестве такой информации могут использоваться, например, данные об отказах или задержках в узлах или каналах связи в ТКС.

В зависимости от используемых способов обработки локальной или глобальной информации (обратной связи) принципы адаптивной маршрутизации можно разбить на три класса:

- централизованная (иерархическая) маршрутизация;
- децентрализованная (распределённая) маршрутизация;
- мульти-агентная (многоадресная) маршрутизация.

Принцип централизованной маршрутизации заключается в том, что каждый узел ТКС сначала передает информацию о своем состоянии (задержки или пропускные способности выходных каналов и т.п.) центральному узлу-маршрутизатору. Затем этот маршрутизатор вычисляет оптимальный маршрут на основе полученной глобальной информации о текущем состоянии и передает его обратно всем узлам ТКС. После этого начинается управляемая передача пакетов от узла-источника к узлу-получателю ТКС по спланированному оптимальному маршруту.

Принцип децентрализованной (распределенной) маршрутизации основывается на обмене локальной информацией между узлами ТКС и использовании этой информации о текущем состоянии узлов и каналов связи ТКС для вычисления оптимального маршрута. По мере вычисления последовательных участков этого маршрута осуществляется распределённо-управляемая передача пакетов от узла-источника к узлу-получателю ТКС.

Принцип мульти-агентной маршрутизации и самоорганизации потоков данных является своеобразным компромиссом между принципами централизованной и децентрализованной маршрутизации. Он основывается на многоадресной и многопотоковой маршрутизации и анализе возможных сетевых конфликтов с целью их предотвращения или разрешения в процессе управляемой передачи пакетов по множеству оптимальных маршрутов от узлов-источников к узлам-получателям глобальной ТКС.

Более подробно этот принцип и конкретные методы мульти-агентной маршрутизации рассмотрены в работах [1-5].

Работа выполнена при поддержке Минпромнауки РФ (проект № 37.029.11.0027), РФФИ (проект № 03-01-00224а) и РГНФ (проект № 03–06–12019в).

Литература

1. Тимофеев А.В. Проблемы и методы адаптивного управления потоками данных в телекоммуникационных системах. – Информатизация и связь, № 1-2, 2003, с. 68-73.
2. Тимофеев А.В. Модели мульти-агентного диалога и информационного управления в глобальных телекоммуникационных сетях. - Труды 10-ой международной конференции "Knowledge-Dialogue-Solution" (16-26 июня, 2003, Варна), 2003, с. 180-186.
3. Тимофеев А.В. Методы высококачественного управления, интеллектуализации и функциональной диагностики автоматических систем. – Мехатроника, автоматизация, управление, 2003, № 2, с. 13-17.
4. Syrtzev A.V., Timofeev A.V. Neural Approach in Multi-Agent Routing for Static Telecommunication Networks. – International Journal "Information Theories and Their Applications", 2003 v. 10, N 2, pp. 167–172.
5. Timofeev A.V. Multi-Agent Information Processing and Adaptive Control in Global Telecommunication and Computer Networks. – International Journal "Information Theories and Their Applications", 2003 v. 10, N 1, pp.54–60.

Информация об авторе

А.В.Тимофеев: Санкт-Петербургский институт информатики и автоматизации, Российской академии наук, tav@iias.spb.su

ЕДНА РЕАЛИЗАЦИЯ НА ХЕШ-ТАБЛИЦА С ОТВОРЕНО АДРЕСИРАНЕ В ПОСЛЕДОВАТЕЛНО РАЗПРЕДЕЛЕНА ПАМЕТ

В. Дянкова, Р. Христова

Анотация: *В статията се разглежда една реализация на хеш-таблица с отворено адресиране в последователно разпределена памет, даваща възможност за намаляване на броя на линейните проби при установяване на факта, че даден елемент не е включен в хеш-таблицата.*

Ключови думи: *хеш-таблица с отворено адресиране, колизия, търсене на елемент, изключване на елемент*

Увод

Извличането на определено парче, или парчета, информация от голям обем предварително запазени данни представлява една фундаментална операция, наречена търсене. Тази операция е неделима част от множество изчислителни задачи. Обикновено целта на търсенето е да се получи достъп до информацията, съдържаща се в елемента и да се извърши последваща обработка. Приложенията на търсенето са широко-мащабни и включват множество различни операции. Общото между тях обаче е, че се нуждаят от добри алгоритми за търсене. Една фундаментална и широко използвана структура от данни, позволяваща прилагането на бързи алгоритми за търсене е хеш-таблицата. Естествено възниква въпросът за реализация използваната структура от данни. При реализирането на една и съща структура могат да бъдат използвани различни подходи, адаптирани към конкретните изисквания за бързодействие, използвана памет и др.

В настоящата статия се разглежда една реализация на хеш-таблица с отворено адресиране в последователно разпределена памет, даваща възможност за намаляване на броя на линейните проби при установяване на факта, че даден елемент не е включен в хеш-таблицата .

Понятия, свързани с хеш-таблици с отворено адресиране.

Таблицата е съвкупност от еднотипни елементи, всеки от които е изграден от ключ (идентификационна част) и тяло (информационна част). Стойността на ключа еднозначно идентифицира конкретен елемент. Стойностите на информационната част не са съществени от гледна точка на организацията и обработката на таблицата, поради което в предложената реализация не е зададен конкретен тип за нея.

В хеш-таблицата достъпът до елемент се осъществява чрез трансформиране на ключа на елемента в неговия адрес, т. е. търсенето може да се определи като изображение *hash*: $K \rightarrow A$, което се нарича хеш-функция. Изображението *hash* дава съответствие на множеството от възможни стойности на ключовото поле *K* в множеството от адреси на паметта *A*, в които физически е поместена самата хеш-таблица.

За представянето на хеш-таблица в последователно-разпределена памет се използва масив, поради което трансформацията на ключ в адрес се свежда до трансформиране на ключ в индекс на масив. Тогава, ако с N бъде означен броят на елементите в масива, то: hash: K $\rightarrow$ {0, 1, 2, … , N-1}

Изборът на добра хеш-функция е гаранция за равномерно разпределение на елементите в хеш-таблицата, но тъй като това няма отношение към целта на настоящата статия, то считайки без ограничение на общността, че ключовете на елементите са естествени числа, ще бъде използвана класическата хеш-функция $hash(k) = k\%N$, където $k \in K$ и k е ключ на елемента.

Тъй като мощността на множеството K е по-голяма от възможния брой на елементите в масива, то е възможно да възникне ситуация, в която два елемента с различни ключове $k_1 \neq k_2$ да претендират за едно и също място в масива, т.е. $hash(k_1) = hash(k_2)$. Такива елементи се наричат синоними, а самото явление – колизия. Проблемът за разрешаване на колизиите има различни решения, но в статията ще бъде разгледан методът на линейното отворено адресиране. При този метод, ако елемент претендира за място, заето от друг елемент, масивът се преглежда последователно за ново свободно място. Последователното търсене би могло да бъде реализирано с функцията $rehash(i) = (i+1)\%N$. Ефектът от разрешаването на колизии при хеш-таблиците с отворено адресиране е т.нар. първично скупчване. Елементите, предизвикали колизиите (с една и съща стойност на hash-функцията) се разполагат последователно след входовете си. Тогава вече и елементите, които трябва да влязат на мястото си, с друга стойност на hash-функцията, ще предизвикат колизии. Случаят, когато за един индекс претендират елементи с различна стойност на hash-функцията се нарича вторично скупчване, а непрекъснатите групи от елементи, които се образуват в таблицата се наричат клъстери.

Основните операции в хеш-таблици и класическите алгоритми за тяхната реализация са:

Търсене на елемент с ключ k - с помощта на хеш-функцията $hash(k) = a_1$ се пресмята индекса a_1 на масива, където би трябвало да бъде намерен този елемент; но ако в нея има елемент с различен ключ, проверката се извършва за позиция $a_2 = rehash(a_1)$. Аналогични действия се извършват и с позиция a_2. Този процес на линейни проби продължава докато:

- бъде намерен елемент с ключ k (успешен край);
- бъде достигнат елемент с нулев ключ (неуспешен край);
- чрез обхождане на цялата таблица бъде установено, че елемент с такъв ключ отсъства (неуспешен край).

Добавяне на елемент с ключ k - ако бъде установено, че елемент с такъв ключ отсъства от таблицата и в нея има свободно място, то той се добавя в таблицата.

Изтриване на елемент с ключ k - след успешен край на операцията търсене на елемент с ключ k и установяване на неговото местоположение в позиция i, се нулира ключа на елемента в тази позиция.

Проблеми за разрешаване.

Още Кнут във фундаменталния си труд [Кнут, 1978] пише: "Многие програмисты свято верит в алгоритмы и очень удивляются, обнаружив, очевидный способ удаления записей рассеянной таблицы не работает.". Прегледа на класическата литература по въпроса показва, че операцията изтриване на елемент от хеш-таблици с отворено адресиране или не се коментира [Амерал, 2001], [Рейнголд, 1980], [Мейер, 1982] или се коментира по някой от следните начини:

- Възможна е реализация на операцията, но тя е много сложна [Амерал, 2001].
- Маркирането на елемента с очевидния метод скъсва веригата на синонимите [Наков, 2002], [Шишков, 1995].
- Въвежда се спецификатор с три състояния: елемента е зает; елемента е свободен, като никога не е бил заеман; елемента е свободен, но преди това е бил зает [Шишков, 1995], [Смит, 2001] и [Седжуик, 2002].
- Вътрешните вериги от синоними се преструктурират, така че изключването на елемент да няма последствия върху алгоритмите за търсене и включване [Шишков, 1995].
- повторно се хешират всички елементи между изтрития елемент и следващата незаета позиция [Седжуик, 2002].

Предложени са две реализации на изтриване на елемент: по очевидния метод в [Азълов, 1995] и чрез повторно хеширане в [Седжуик, 2002].

Така дадената дефиниция на операцията търсене и коментирания по-горе проблем при операцията изтриване се пораждат следните въпроси:

2.1. Как се отразява изтриването на елемент от веригата линейни проби върху търсенето на елемент, влизащ в състава на тази верига?

2.2. Достатъчно основание ли е достигането на нулев ключ във веригата линейни проби при търсене на елемент, за да се твърди, че търсеният елемент не е в таблицата?

2.3. Необходимо ли е обхождане на таблицата до достигане на празен елемент (който никога не е бил запълван), за да бъде установено, че търсеният елемент отсъства?

Един пример

Като илюстрация на повдигнатите въпроси може да бъде разгледан следния пример: в хеш-таблица с размер 13 се разполагат елементи с ключове

- *14 (hash(14)=14%13=1);*
- *16 (hash(16)=16%13=3);*
- *29 (hash(29)=29%13=3, rehash(3)=4%13=4);*
- *55 (hash(55)=55%13=3, rehash(3)=4%13=4, rehash(4)=5%13=5);*
- *21 (hash(21)=21%13=8);*
- *35 (hash(35)=35%13=9);*
- *49 (hash(49)=49%13=10);*
- *50 (hash(50)=50%13=11).*

0	1	2	3	4	5	6	7	8	9	10	11	12
0	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**21**	**35**	**49**	**50**	**0**

При прилагане на разпространения в литературата алгоритъм за търсене за елементи с ключове 14, 20, 55, 42, 48 се получават следните резултати и възникват следните въпроси:

- за 14: чрез прилагане на хеш-функцията се получава индекс 1, където се намира търсения ключ, т.е. търсенето завършва успешно;
- за 20: чрез прилагане на хеш-функцията се получава индекс 7, където стои ключ 0, т. е. мястото е празно и търсенето завършва неуспешно;
- за 55: чрез прилагане на хеш-функцията се получава индекс 3, където стои ключ 16≠55. Следвайки класическия алгоритъм (извършват се линейните проби до достигане на елемент с търсения ключ или до празно място) след прилагане на метода на линейните проби, се достига до ключа 55 и търсенето завършва успешно;
- за 42: чрез прилагане на хеш-функцията се получава индекс 3, където стои ключ 16≠42. Прилагайки метода на линейните проби, се достига до ключ 0, т. е. до празно място и търсенето завършва неуспешно;
- за 48: чрез прилагане на хеш-функцията се получава индекс 9, където стои ключ 35≠48. Прилагайки метода на линейните проби, се достига до ключ 0, т. е. до празно място и търсенето завършва неуспешно. В този случай би могъл да бъде поставен следния въпрос: не може ли още при получаването на индекс 9 да направим извода за неуспешен изход на търсенето, тъй като както се забелязва от процеса на разполагане на елементите в таблицата в позиция 9 не е възниквала колизия. Тогава няма как да се очаква търсения ключ да се появи във веригата линейни проби за тази позиция (въпрос 2.3).

За илюстрация на повдигнатите въпроси 2.1. и 2.2. може да бъде разгледано търсенето на елемент с ключ 55 след изтриването на елемент с ключ 29. Разположението на елементите след операцията изтриване е следното:

0	1	2	3	4	5	6	7	8	9	10	11	12
0	**14**	**0**	**16**	**0**	**55**	**0**	**0**	**21**	**35**	**49**	**50**	**0**

Тогава при прилагане на алгоритъма за търсене на елемент с ключ 55 и обхождайки индексите от линейните проби се достига до индекс 4, където стои ключ 0, т. е. мястото е празно и съгласно класическия алгоритъм (търсене до откриване на ключа или празно място) търсенето би трябвало да завърши неуспешно, което противоречи на действителното разположение на елементите в таблицата. Противоречието възниква поради факта, че в позиция 4 е възниквала колизия и следователно ако там не фигурира търсения ключ, би трябвало да се продължи с линейните проби до достигането на търсения ключ (успешен край) или до позиция, в която не е възниквала колизия (неуспешен край).

Решение на проблема

Основния проблем, на който се предлага решение в статията е реализацията на алгоритъм за търсене на елемент в хеш-таблица с отворено адресиране, при който се намалява броя на линейните проби при установяване на факта, че даден елемент не е включен в хеш-таблицата, а изключването на елемент няма неприятни последствия върху алгоритмите за търсене и включване на елемент във вътрешните вериги на синонимите при разрешаване на колизии. Освен това не се налага използването на процес "събирач на боклук", който е много бавен.

Разрешението на този проблем, което се предлага в настоящата статия е за всяка позиция да се пази информация дали на съответното място е възниквала колизия при добавянето на елементи. Това води до идеята за алтернативен масив *ph* от булеви стойности:

$$ph[i] = \begin{cases} true, & ако \quad на \quad i-то \quad място \quad е \quad имало \quad колизия \\ false & в \quad противен \quad случай \end{cases}$$

Тогава разполагането на разгледаните по-горе елементи ще предизвиква следните промени в двата масива:

- *14 (hash(14)=14%13=1):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *16 (hash(16)=16%13=3):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *29 (hash(29)=29%13=3; rehash(3)=4%13=4);*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *55 (hash(55)=55%13=3, rehash(3)=4%13=4, rehash(4)=5%13=5):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*true*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *21 (hash(21)=21%13=8):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**21**	**0**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*true*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *35 (hash(35)=35%13=9):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**21**	**35**	**0**	**0**	**0**
ph	*false*	*false*	*false*	*true*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *49 (hash(49)=49%13=10):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**21**	**35**	**49**	**0**	**0**
ph	*false*	*false*	*false*	*true*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

- *50 (hash(50)=50%13=11):*

	0	1	2	3	4	5	6	7	8	9	10	11	12
t	**0**	**14**	**0**	**16**	**29**	**55**	**0**	**0**	**21**	**35**	**49**	**50**	**0**
ph	*false*	*false*	*false*	*true*	*true*	*false*	*false*	*false*	*false*	*false*	*false*	*false*	*false*

При това състояние на масивите търсенето на елемент с ключ 18 ще доведе до прилагане на хеш-функцията над ключа 18 – *hash(48)=48%13=9*. Тъй като на това място не стои елемент с търсения ключ и $ph[i] = false$, то не е необходимо прилагането на линейните проби и може да бъде направен извод за неуспешен край на търсенето.

Програмна реализация

Следва една реализация на C++I на предложеното решение:

```
const int nilkey=0;
template<class T>
struct element {int key; T info;};
template<class T>
class hashtable
{ private:
   int tabsize;
   int free;
   element* t;
   bool* ph;
  public:
   hashtable();
   hashtable (int n);
   bool is_full();
   int search (int k);
   void insert (element e);
   void del (int k);  };
template<class T>
hashtable<T>::hashtable()
{ tabsize = 0; free=0; }
template<class T>
hashtable<T>::hashtable(int n)
{ tabsize = n; free=n;
  t = new element[tabsize];
  ph = new bool[tabsize];
  for (int i=0; i < tabsize; i++)
      { t[i].key=nilkey; ph[i]=false; }
}
int h(int k)  { return k%tabsize; }
int r(int i)  { return (i+1)%tabsize; }
template<class T>
bool hashtable<T>::is_full()
{ return free==0; }
template<class T>
int hashtable<T>::search( int k)
{ bool b=false; int i=h(k); int j=i;
  while ( t[i].key!=k && ph[i] && !b )
        { i=r(i); b = i==j; }
  if (t[i].key==k) return i;
  else return –1;
}
template<class T>
void hashtable<T>::insert( element E)
{ int i=search(E.key);
  if ( i<0 && !is_full() )
```

```
    { i=h( E.key);
      while ( t[i]!=nilkey ) { ph[i]=true; i=r(i); }
      t[i]=E; free--;
}  }
template<class T>
void hashtable<T>::del( int k)
{ int n=search(k);
  if (n>=0) { t[n].key=nilkey; free++; }
}
```

Заключение

Броят на линейните проби при търсене на елемент в хеш-таблица с отворено адресиране зависи от:

1. съотношението α на заетите елементи спрямо всички елементи в таблицата, като при една разредена таблица(α е малко) очакваме повечето търсения да открият празна позиция само след няколко проби; при една почти пълна таблица(α е близко до 1) търсенето може да изисква голям брой проби;

2. начинът, по който при вмъкването си елементите образуват непрекъснати групи (клъстери) в таблицата, като наблюденията показват, че усреднения брой проби при неуспешно търсене е пропорционален на квадратите на дължините на клъстерите, а успешните търсения винаги по-евтини от неуспешните.

Предлаганата реализация на хеш-таблица с отворено адресиране чрез масива *ph* фактически намалява съществено дължината на клъстерите и по този начин намалява броят на пробите при възникването на колизии, особено ако търсения елемент не е от хеш-таблицата.

Предлаганата реализация на хеш-таблица в последователно разпределена памет може да се използва с произволни хеш-функции и произволен начин за обработване на колизиите, като във всеки от тези случаи съществено се намалява броят на пробите при търсене на елемент. Операциите добавяне на елемент и изтриване на елемент пряко или косвено използват търсене на елемент и скоростта на тяхното изпълнение също се подобрява. Допълнителна полза е ,че изтриването на елемент работи коректно без да се налага използването на процес “събирач на боклук”.

Литература

[Азълов, 1995] П.Азълов. Програмиране. Основен курс. Изд. АСИО, София, 1995.

[Амерал, 2001] Л.Амерал. Алгоритми и структури от данни в С++. ИК СОФТЕХ, 2001.

[Кнут, 1978] Д.Кнут. Искусство програмирования для ЭВМ, т.3. Сортировка и поиск. Изд. Мир, Москва, 1978.

[Мейер, 1982] Б.Мейер, К.Бодуэн. Методы программирования. т.2. Изд. Мир, Москва, 1982.

[Наков, 2002] П.Наков, П.Добриков. Програмиране=++Алгоритми; Изд.TopTeam Co., София,2002

[Седжуик, 2002] Р.Седжуик. Алгоритми на С, ч.4. Търсене. Изд. СофтПрес, 2002

[Смит, 2001] Т.М.Смит. Принципи и методи на програмирането с PASCAL. Изд. Техника, София, 2001

[Рейнголд, 1980] Э.Рейнголд,Ю.Нивергельт,Н.Део. Комбинаторные алгоритмы. Теория и практика. Изд.Мир, М., 1980

[Шишков, 1995] Д.Шишков и др. Структури от данни. Изд.Интеграл, Добрич, 1995

Информация за авторите

Валентина Спасова Дянкова – ШУ “Епископ К.Преславски”, ФМИ; ул.”Университетски” 115, Шумен-9712; e-mail: vspasova@yahoo.com

Росица Христова - ШУ "Епископ К. Преславски", ФМИ; ул.”Университетска” 115, Шумен-9712;

e-mail: r.hristova@fmi.shu-bg.net

ОРГАНИЗАЦИЯ НА ДИАЛОГА С ПОТРЕБИТЕЛЯ В КОМПЛЕКС ФОИ

Красимира Иванова, Илия Митов, Красимир Марков

Резюме: *В настоящата статия се представя система от четири базови информационни структури, предназначена за изграждане програмни системи за информационно обслужване.*

Въведение

Развитието на програмните системи за информационно обслужване достигна като че ли своята връхна точка след бума на базите от данни. Днес трудно можем да се представим реална промишлена или търговска дейност (бизнес), при която да не се използват компютри за едни или други цели. Програмните системи за информационно обслужване буквално заляха света. Този процес продължава с нестихваща интензивност.

На фона на това експоненциално нарастване се забелязва и нарастване на интереса към все по-нови възможности за изграждане на програмното осигуряване, чрез което се реализират приложенията. В теоретичен план се развиха редица направления за автоматизация на проектирането и създаването на приложни системи. Вече има редица продукти, които подпомагат по един или друг начин проектантите и реализаторите на системи за информационно обслужване.

В търсенето на все по-нови възможности, обаче, често се пропуска систематичността и информационното моделиране, които следва да предхождат методологически новите разработки. Ето защо тези фундаментали елементи на всяка приложна реализация придобиват все по-голяма важност и значимост.

Изграждането на прогностични информационни модели и систематичното им реализиране е базов подход, който авторите на настоящата статия постоянно следват. Ползата от това е, че с ограничени човешки и икономически ресурси се постига висок реализационен и практически значим резултат.

Една въвеждаща в методологията статия е [Марков и др., 1994]. Рамките на една статия ограничават възможностите за по-пълно представяне на цялата методология. Тук ще продължим нейното разглеждане като представим системата от информационни структури, която авторите ползват при изграждането на приложни програмни системи. През последните петнадесет години тези системи придобиха популярност в България под името “Комплекс ФОИ”. Днес Комплекс ФОИ се радва на над хиляда инсталации в цяла България. Една от много важните методологични черти на комплекса е организацията на диалога с крайния потребител, позволяваща бързо и без особени усилия да се усвои работата с комплекса без особена предварителна компютърна подготовка и знания, но с добро познаване на съответната практическа област.

Работата на крайния потребител със системите от Комплекс ФОИ е пределно проста. Необходимо е той да привикне да работи с четири основни диалогови структури: *меню, йерархия, табло, текстов файл*.

Можем да отбележим, че тази система от информационни структури се оказа функционално пълна – чрез нея бяха реализирани всички системи на Комплекс ФОИ, без да е необходимо добавянето на нови структури. Отсъствието, например на иконни изображения, не се оказа недостатък на интерфейса. Напротив, систематичното изграждане и подреждане на екранните форми, базирано на скрит от потребителя сценарен подход, позволи да се поддържа работата на потребителя в съответствие с правилата, по които той следва да извършва своята ежедневна професионална дейност.

Някои от представените по-долу структури са добре познати и по тази причина те ще бъдат само кратко споменати, а вниманието ни ще се съсредоточи на оригиналните елементи.

Работа с менюта

Менюто представлява удобен апарат за избор от група еднородни помежду си елементи. Обикновено това са имена на функции, в някои случаи - възможност за задаване на стойности на параметри, а в други – данни, които са определени твърдо в системата или потребителят сам разширява в процеса на работата си.

Менюто представлява вертикално подредени текстови или числови елементи. Изборът на даден елемент от менюто става чрез натискане на стрелките "нагоре" и "надолу", при което избраният елемент се осветява с различен цвят. С натискане на "Enter" се предизвиква активиране на този ред от менюто. Натискането на "Esc" предизвиква излизане от менюто и връщане в предходна операция.

Особена възможност е предоставена чрез т.н. ***множествен избор***, когато потребителят има право да избере едновременно няколко (незадължително последователни) елемента.

Допустимо е в някои случаи крайният потребител да добавя ***нови елементи в менюто***, когато това е методологически оправдано – разширяване състава на дадена номенклатура, формиране на списъци от индивидуални настройки на справки, отчети и др.

Накрая ще отбележим, че има възможност крайният потребител временно да забранява избора на отделни елементи в дадено меню. Обикновено по този начин потребителите се предпазват от неволно изпълнение на функции, които предполагат внимание и концентрация при изпълнението им, например възстановяване на по-ранна версия на архивите, която ще замени (унищожи) текущата.

Работа с дървовидна организация на информацията

Чрез дърво просто и лесно се изгражда описание на големи множества от обекти, които са организирани като йерархия. Дървовидната структура се използва например: при описание и работа със сметкоплана на една фирма, при задаване структурата и съответната подчиненост на звената и кадрите ѝ, при описването на материалите в различните складове и др.

Дървовидните структури в системите от Комплекс ФОИ са достъпни с помощта на специална функция от главното меню на съответните системи. След влизане в дървовидната структура на първия ред в горната част на екрана се изписва кодът на текущия възел, предхождан от кодовете на възлите, през които се минава за да се стигне до възела. На следващия ред е името на текущия възел.

Когато възелът има подвъзли, техните кодове и имена се изписват в подекран. Курсорът обикновено е позициониран на първия подвъзел.

Движение по дървовидната стуктура

Ако желаем да слизаме надолу по дървото, с движение на стрелките "надолу" и "нагоре", "PgUp" и "PgDn", "Home" и "End", избираме желания подвъзел, който искаме да направим активен и натискаме "Enter". Ако желаем да се върнем към предходния възел натискаме "Esc".

Функции за опериране с дървовидни структури

При работата с дървовидните структури, освен функциите за движение по нивата на структурата, може да се използват и следните функции:

- **поправка на името на възела:** Функцията позволява да променяме номера и името на текущия възел.
- **добавяне на подвъзел:** Тази функция позволява да се добави подвъзел на текущия в момента възел. Текущият възел става предходник на този, който се създава.
- **вмъкване на подвъзел:** Чрез тази функция се осъществява вмъкване на подвъзел. Изпълнението ѝ е аналогично на горната, с тази разлика, че вмъкнатият подвъзел се разполага над осветения в списъка вляво на екрана.
- **преместване на възел:** Смяната на местоположението на даден възел в дървовидната структура, а също и някои по-сложни операции по обединението на елементите на различни възли, се осъществяват с помощта на тази функция.
- **изтриване на възел:** При изтриването на възела системата задава контролни въпроси за потвърждаване на изпълнението на операцията или извежда съобщения при невъзможност да бъде изтрит възела (има подвъзли или съдържа съществени данни за конкретната система). След изтриването на текущия възел системата автоматично прави текущ възела, който е съдържал този възел.
- **разпечатка и сортиране на възела:** Позволява текстово извеждане на съдържанието на частта от дървото, която е подчинена на текущия възел. Извеждането може да стане в съответствие с подчинеността в дървовидната структура или сортирано по азбучен ред. От този ред от менюто могат да се изпълнят и функции, които позволяват да се сортират въведените подвъзли “по номера” или “по

имена". Сортировката засяга само пряко подчинените подвъзли, а не и техните наследници (там се запазва зададената подредба на елементите).

- **инициализация на възел:** Тази функция се използва за указване на системата кои възли (без подвъзли) стават активни при изпълнението на функциите за работа с дървото, както и за задаване на специфични настройки на възела и начални стойности на параметрите му;
- **специфични функции:** касаят конкретната система, използваща дървовидна структура.

Функциите, включени в менюто за работа с дървото, се отнасят винаги за текущия възел.

Когато възелът няма подвъзли, вдясно на екрана директно се появява прозорецът с менюто на функциите към възела. Когато възелът има подвъзли те се изобразяват в лявата част на екрана, а това меню се показва/скрива с помощта на клавиша за табулация **"Тав"**.

Работа с табло

Апаратът за работа с табла е изграден в съответствие с идеите на CellPro подхода (клетъчно-ориентирано програмиране) [Markov et al., 1995] и сценарно-базирано подпомагане на последователността на изпълнение на действията на крайния потребител [Markov et al., 1999].

Таблото представлява съвкупност от клетки (полета), разположени на екрана. Всяка клетка има самостоятелно значение и работата с нея е относително автономна. За да може да се въвежда информация в дадена клетка или да се активира нейната функция е необходимо тя да бъде избрана, след което се натиска "Enter".

Клетки

При изграждането на информационните системи всяка клетка се определя от своя номер и от описателя си. В описателя на клетката се съдържа информация за графичните характеристики на клетката, вида на клетката, координатите на клетката, зависимостта на клетката от другите клетки на таблицата, връзката й с данни от информационната база.

Най-общо клетките се делят на две групи - в които потребителят се допуска или не се допуска да работи.

Клетките могат да бъдат от следните типове:

- s, z: в такива клетки се въвеждат текстови данни;
- i, x, r, y: данните, които се въвеждат са цели или реални числа;
- d: клетки, съдържащи дати;
- t: клетката от този тип се явява вход към друга таблица;
- p: от клетката може да се извиква програма, чиито резултати могат да се показват (ако е необходимо) или в самата клетка, в някоя друга клетка или в рамките на целия екран. След приключването на програмата системата продължава да работи в таблицата;
- f: това е клетка в която потребителят може да стартира определена функция (предварително създадена в системата), която връща в клетката резултата от изпълнението на дефинираната операция;
- l: само за четене, без да може да се измени архива с данни.

Координатите на една клетка се задават с наредената петорка (T,X,L,Y,H), където:

- T - показва как трябва да възприемаме координатите X и Y – като абсолютни стойности от началото на екрана или относително спрямо предходната клетка;
- X - начална стойност на левия горен ъгъл на клетката по оста X;
- L - ширина на клетката;
- Y - начална стойност на левия горен ъгъл на клетката по оста Y;
- H - височина на клетката.

При относително отчитане на координатите X и Y могат да бъдат и отрицателни.

В общия случай документите(таблиците) не могат да се изобразят изцяло на екрана на терминала поради неговите ограничени размери.

Екранът на терминала се разглежда като един виртуален екран, който се мести по таблицата но не с една клетка наляво или надясно и с един ред надолу или нагоре, а с цял екран в съответната посока. Това се налага поради факта, че в общия случай изобщо може да не са определени редовете и стълбовете в документа. При това положение понякога е неудобно да се работи защото част от информацията, например заглавието на таблицата, се намира в някакъв друг екран. Този проблем е разрешен чрез определяне на зависимости между клетките на таблицата.

Има два вида зависимости между клетките:

- зависимост по връзка: пряка или косвена;
- зависимост по месторазположение: твърда връзка до/под или свободна връзка - по избор на системата в рамките на екрана.

В описателя на клетката се съдържат данни за това, къде в архива за данни тази клетка ще занесе въведената в нея информация или съответно - откъде ще получи такава. Тези данни се съдържат в наредената тройка - /F,N,K/, където F - име на архив или файл с данни, N - показва номера на областта, а K - номера на обекта. Тези данни могат да се зададат при задаването на описанието на самата клетка, но могат да се конкретизират при извикването на съответното табло, използващо клетките. Апаратът за задаване на описателя позволява гъвкаво изместване по областите или обектите на координатното пространство [Марков, 2004] или твърдо фиксиране на определена характеристика.

Движение по таблото

Движение по клетките в таблото става със стрелките "наляво" и "надясно". Когато таблото (или част от него) има таблична структура по колоните на таблиците можем да се движим, използвайки стрелките "надолу" и "нагоре".

В повечето случаи размерите на екрана са недостатъчни за да се изобрази цялата информация от таблото (например таблица с няколко хиляди реда). Можем да си мислим, че изображението на екрана е един прозорец, през който се наблюдава част от таблото. Този прозорец може да се премества по таблото нагоре и надолу с използване на клавишите "PgDn" (страница надолу) и"PgUp" (страница нагоре) - по аналогия с прелистването на книга. Преминаването към точно определена страница се активира с помощта на комбинацията клавиши "Alt"+"J" (скок/преход към) и задаване на номера на страницата.

Има по-сложни структури, в които един елемент се разполага на няколко екрана. Тогава преминаването от една структура към следващата или предходната, става чрез натискане на “Ctrl”+”PgDn” или “Ctrl”+”PgUp”. При движението по клетките на таблото се намираме в режим на наблюдение. Ако искаме да редактираме или активираме функция в някоя клетка трябва да натиснем "Enter", което предизвиква превключване на режима.

Връщането в предишния режим става след завършването на изпълнението на функцията, назначена на клетката, или с натискане на "Enter" или "Esc", когато редактираме.

Към всяка клетка е осигурена възможност за получаване на коментарна информация с натискане на "Alt"+"H".

Редактиране в клетка

Редактирането в клетката става чрез използването на клавишите, с които сме свикнали да работим в текстовите редактори: "стрелките", "Home", "End", "BcSp", "Del", "Ins". Приключването на редактирането става чрез натискане на:

- "Esc": предизвиква излизане от режима на редактиране и възстановяване на съдържанието на клетката каквото е било преди започване на редактирането;
- "Enter": приключва редактирането и съхранява направените промени в клетката.

Калкулатор

При натискане на комбинацията "Alt"+"K" се включва вграден в системата калкулатор. Работата с калкулатора се завършва по два начина:

- чрез натискане на "Esc", при което се излиза от калкулатора и се загубва текущо полученият резултат;
- чрез натискане на "Enter", при което текущият резултат се запомня в клетката, в която е бил курсорът преди активирането на калкулатора, ако типът на клетката позволява това.

Изтриване на съдържимото в клетката

В случай, че дадена клетка не е свързана с по-особена обработка на информация, с натискане на комбинацията от клавиши ”Alt”+”D” се изтрива съдържанието на клетката. В някои случаи изпълнението на тази операция предизвиква изтриване на голяма част от информация, свързана смислово с тази клетка –

например, активирането на тази операция върху клетките “системно име на архив” предизвиква изтриването на потребителските архиви, свързани със съответния ред от таблицата. На такива по-отговорни места системата иска потвърждение за желанието да се изпълни операцията по изтриване.

Копиране от горната клетка

В някои от клетките за редактиране е възможно използването на комбинацията "Alt"+"C", която предизвиква пренасяне (копиране) на съдържанието на клетката, разположена на екрана над текущата клетка, в текущата. Това е полезно при работа с таблици.

В клетки от табло, които смислово са свързани с по-сложни информационни структури, например, цели документи, архиви или отделни структури от архиви, комбинацията от клавиши "Alt"+"C" служи за копиране на тези структури една в друга, като в този случай за по-голяма гъвкавост системата изисква да се укаже откъде ще се копира.

Маркиране на група и изпълнение на функции върху нея

При работа с таблици понякога се налага извършването на определени операции върху група редове от таблицата. Видът на функцията се определя от колоната, в която е извършена маркировката и е смислово свързана с действието на клетките от тази колона.

Работа с текстови файлове

Системите от Комплекс ФОИ извеждат основните резултати в текстови файлове. Това е удобна форма на представяне на информацията, поради възможността за отпечатване на хартиен носител от една страна и за по-компактно изобразяване на резултатите на екрана, от друга страна.

Изходящата информация се подготвя по два основни начина:
- чрез апарат за цялостно изграждане на изходящата информация;
- на базата на предварително подготвени макетни файлове.

При първия начин системите сами се грижат за вида на изходящата информация, като потребителят може (в някои справки) да настройва само вида, съдържанието и големината на таблиците, извеждани от дадената справка.

Вторият начин се обслужва от генератора на отчети FgR към Комплекс ФОИ, който позволява създаването на бланкови документи на базата на информация от архивите на системите от Комплекса. Всеки отчет представлява макетен текст, в който, на избрани от потребителя места, генераторът вмъква данни от архивите. За целта се ползва прост език за описание на препратките. Символът “#” се използва за заграждане на ключовите изрази, които системата при изпълнението си заменя с текущата потребителска информация.

Системите са снабдени с програми, позволяващи изведената във файла информация да бъде наблюдавана и отпечатвана в удобен за потребителя вид. Текстовите файлове, които се формират от системите, обикновено се маркират с разширение ".TXT".

Съществува апарат, позволяващ да се укаже кои програми за извод на резултата ще използва потребителят, включително и други редактори или електронни таблици, които не принадлежат към Комплекса.

При инсталирането на Комплекса автоматично се задават две програми за извод на резултатите, които са част от Комплекса – FED_D: за работа в DOS-среда и FED_W: за работа в среда WINDOWS.

Добавянето на други програми или изключването на някоя от наличните става в специален прозорец за общи настройки към Комплекса.

Заключение

В настоящата статия бяха представени основните информационни структури, с чиято помощ се изгражда диалога с потребителя в програмните системи от Комплекс ФОИ.

Простотата за възприемане и лекотата при работа с тях са важно предимство, което ги направи предпочитани в нашата работа. Въпреки това, следва да отбележим, че дори и с такива средства е възможно да се построят претрупани и объркани интерфейсни процедури. Ето защо, важна особеност на възприетия подход е системата от ергономични правила, които се следват при построяване на диалога с потребителя.

На първо място е отчитането на психологическите особености на човека за концентрация на вниманието при работа с множество информационни обекти. На тази тема следва да се посвети отделна публикация и затова тук ще се ограничим само с два примера.

Добре известно е, че капацитетът на кратковременната памет у човека не е голям. От десетилетия е идеята, че около седем обекта (плюс-минус два) могат да се поддържат активни в нея [Miller, 1956]. Ето защо, диалоговите информационни структури се организират на принципа на групиране така, че на потребителя да не се налага да работи с повече от два-три информационни обекта едновременно.

Едно от следствията от ограниченията на кратковременната памет на човека е същественото облекчение, което той получава, когато изчезне необходимостта от по-нататъшно съхраняване на информацията. Това предизвиква силно желание да се завърши задачата, да се намали натоварването на паметта и да се получи кратка почивка. Това е т.н. "затваряне" на текущата задача [Shneiderman, 1980]. В диалоговите информационни структури, които бяха представени в настоящата статия, възможността за разпределяне на работата в последователност от завършени стъпки е водещ принцип. Това най-силно се вижда в таблата, където всяка стъпка е ограничена в рамките е една единствена клетка и приключването на работата в клетката (затварянето й!) е именно онзи желан от потребителя миг на кратък отдих на вниманието, за да премине с нов психологически импулс към работата в следващата по сценарий клетка, където системата вече автоматично се е позиционирала.

Накрая следва да отбележим, че представената в настоящата работа система от диалогови информационни структури не следва да се разглежда като претенция за единственост и незаменимост. Точно обратното – тази система е само една от многообразието, което информатиката вече е предложила. Безспорно е, че всяка една подобна система би била много полезна, ако се прилага систематично и ергономично.

Благодарности

Настоящата работа частично е финансирана по проект ITHEA-XXI на Института по информационни теории и приложения FOI ITHEA.

Литература

[Марков и др., 1994] Кр.Марков, Кр.Иванова, И.Митов, Й.Икономов, В.Василев. *Информационно моделиране на фирмени дейности.* ИЗВЕСТИЯ: Списание на Икономическия Университет Варна, бр. 3-4, 1994, стр.119-127.

[Markov et al., 1995] K.Markov, K.Ivanova, I.Mitov. *Cell Oriented Programming.* International Journal "Information Theories & Applications"; Vol.3, No.:1; 1995; pp.3-8.

[Markov et al., 1999] K.Markov, K.Ivanova, I.Mitov. *Multiagent Information Service Based on Scripts*. Journal "Штучний інтелект", No.:2, 1999, pp. 129-135.

[Марков, 2004] К. Марков, Координатно базирана физическа организация за компютърно представяне на информационни пространства. Настоящият сборник.

[Miller, 1956] Miller G.A. *The magical number seven, plus or minus two: Some limits on our capacity for procwssing information.* Psychological Review, 63, 1956, pp.81-97.

[Shneiderman, 1980] Shneiderman B. *Software psychology: Human factors in computer and information systems*. Winthrop Publishers Inc., Cambridge, Massachusetts, 1980.

Информация за авторите

Красимира Иванова – Институт по математика и информатика БАН. e-mail: foi@nlcv.net

Илия Митов - Институт по информационни теории и приложения FOI ITHEA. - e-mail: foi@nlcv.net

Красимир Марков - Институт по информационни теории и приложения FOI ITHEA. Институт по математика и информатика БАН. e-mail: foi@nlcv.net

РАЗШИРЕНА ИНФОРМАЦИОННА СИСТЕМА ЗА РЪКОВОДИТЕЛЯ (EXTENDED EIS)

Тодорка Ковачева

***Резюме**: В настоящото изложение се разглежда нов клас информационна система за ръководителя, базирана на принципа на самоорганизация в управлението и изградена въз основа на модулно моделиране. Тя е мултифункционална и мултидисциплинарна. Представени са нейните структурни компоненти и е направена кратка характеристика на отделните модули.*

***Ключови думи**: разширена информационна система за ръководителя, еволюционно управление, управление на конфликти, модулно моделиране, склад от данни, самоорганизация, невролингвистично програмиране .*

Увод – необходимост от нов клас системи

Управлението на предприятията се осъществява от личности, заемащи ръководни постове, като на най-високите нива се позиционират така наречените топ-мениджъри. Те поемат цялата отговорност за осъществяването на управленския процес от момента на целеполагане, до конкретната реализация на поставената за изпълнение фирмена цел. Принудени са непрекъснато да вземат решения, някои от които в критични за предприятието ситуации.

Топ-мениджърите боравят с големи обеми от информация, породени от постоянно циркулиращите вътрешни и външни информационни потоци. Те контролират осъществяването на редица дейности като целеполагане и планиране в предприятието; интегриране на фирмените ресурси; подбор и мотивация на персонала; изпълнение на плановите задачи; управление на финансовите средства и други, свързани с управленския процес. Необходимо е да са максимално информирани за действията и намеренията на конкурентите, за пазарните тенденции, както и да следят технологичното развитие на обществото, с оглед своевременното внедряване на новите технологии в сферата на бизнеса. Всяка от тези дейности предполага наличието на тесни специалисти в съответната област. Топ-мениджърите трябва да притежават всички тези способности и да могат да ги изпълняват. В противен случай не биха могли да осъществят ефективен контрол и управление.

Мениджърите са хора, които не разполагат със свободно време. Огромните обеми информация, с които оперират, не могат да бъдат обхванати, осмислени и своевременно разпределени. Човешкият мозък не е в състояние да се справи с тях на това ниво от еволюционното си развитие. Съществува вероятност значима и необходима за предприятието информация да бъде пропусната, което е предпоставка за наличие на нереализирани възможности.

Всичко това повишава напрежението в работата на ръководителя и води до постоянно високи нива на хроничен стрес. Понижава се качеството на фирменото управление. Последното зависи преди всичко от хората, а те са биохимични системи и следователно се влияят от протичащите в тях биохимични процеси. Стресът намалява работособността и бдителността на мениджърите. Броят на инфарктите, инсултите и другите заболявания, свързани с напрежението, се повишава. Топ-мениджърите нямат време за личен живот и не могат да си позволят пълноценна почивка. Всяко откъсване от работата ще изисква от тях да се справят с голям обем натрупана във времето информация, при това паралелно с тази, която е актуална в момента.

Въз основа на изложеното до тук може да се изрази становището, че всяка иновация, която се прави в областта на управлението, следва да бъде насочена към подпомагане работата на ръководителя в посока към повишаване на неговото свободно време и понижаване нивата на стрес в ежедневието. Това налага да се създадат нов клас системи, които да способстват за постигане на фирмена самоорганизация и които да са в пряка помощ на топ-мениджърите. Те трябва да отговарят максимално точно на потребностите на управляващите и да съответстват на изпълняваните от тях дейности. Такава именно система се явява предлаганата от мен разширена информационна система за ръководителя (Extended Executive Information System – Extended EIS).

Обща характеристика на разширената информационна система за ръководителя

Осъществявайки своята дейност, мениджърите ежедневно вземат решения от оперативен, тактически и стратегически характер. Тези решения са толкова по-добри, колкото по-точно съответстват на желанието на субекта на управление и водят до неговото реализиране при минимален разход на ресурси и за възможно най-кратко време. По такъв начин се насочва и контролира развитието на предприятието и се създават условия за постигане на фирмените цели.

Изборът на мениджърът, представен под формата на управленско решение, е процес, който насочва поведението на обекта на управление в желаната посока и води до промяна на неговото текущо състояние в ново, по-съвършено бъдещо състояние, като го подобрява, в резултат на което обекта еволюира. Ръководителите вземат решения през целия процес на управлението. Последният се представя като съвкупност от общоуправленски функции [1]. Те са целеполагане, планиране, организиране, мотивиране, контролиране и регулиране. Нито една от тези фунцкии не може да съществува без останалите и нито една от тях не е по-значима от другите. Те следва да се интегрират в единна система на управление. Това ще способства за адекватно подбиране на помощни средства, методи и технологии, водещи в своята съвкупност до реализация на поставените цели.

Разширената информационна система за ръководителя следва да покрива всички управленски етапи и да подпомага мениджърите при вземането на решения от различен характер. Обхваща етапите целеполагане, познаване на обекта на управление, моделиране, създаване на идеален модел, оценка на модела, изработване на стратегия за постигане на целта, прилагане на приетата стратегия, отстраняване на допуснатите грешки, обновяване и повторно прилагане на стратегията, оценка на резултата и повтаряне на цикъла до пълното му завършване, стабилизиране и последващо развитие на системата:

1.Точно и ясно определяне на целта на управлението. Целеполагането е първата фаза на управленския процес. То определя посоката, в която ще се развива предприятието. Целта трябва да бъде конкретно и точно формулира и да не съдържа в себе си противоречия. Нейното поставяне, обаче, не е достатъчно условие за реализацията и. Необходимо е да се проведат редица мероприятия, най-важното от които е оценката на целта. Тя включва определяне степента на нейната постижимост, полезност и непротиворечивост, необходимото време за осъществяването и, и други. Това е важно, тъй като съществува реална възможност поставената за изпълнение цел да бъде невъзможна за реализация поради неотчитане на противоречащи и фактори.

Extended EIS на този етап следва да поддържа ***модул за оценка на целта***. Той трябва да включва критерии, по които да се прави това и да има връзка с останалите модули. Част от тези критерии могат да се заимстват от невролингвистичното програмиране. Последното представлява ново направление в психологията, обособило се като самостоятелна наука. Помага да се даде отговор на въпроси като: "Постижима ли е целта?", "Реализацията и само от собствени на предприятието ресурси ли зависи или се разчита и на външни фактори?", Какви са ползите и недостатъците от нейното осъществяване?", "Противоречи ли тя на други фирмени цели?" и така нататък. Съзнателното и точно формулиране на целите повишава шанса за превръщането на желанията в съответни действия, водещи към тяхната реализация [2]. НЛП се занимава не само с целеполагането, но и с целия процес на управлението, поради което считам, че следва да намери по-голямо приложение в икономическата теория и практика.

2.Запознаване с обекта на управление. Всеки обект може да се представи като система – "съвкупност от елементи и връзките между тях и с околната среда, взети заедно с техните свойства и свойствата на връзките, която функционира като едно цяло за постигане на целта [3]". На този етап следва да се събере максимален обем информация, която го характеризира, за неговите елементи и поделементи, с техните свойства и връзките помежду им. Това са базисните проучвания за системата, с които се установяват определени зависимости, тенденции в поведението, принципи на действие и други. На този етап се допускат и едни от най-сериозните грешки, поради непълно обхващане и подценяване на значими фактори или включване на излишни такива, което утежнява работата и може да доведе до нереалистични резултати. Нивото на детайлизация трябва да се определи много внимателно. Софтуерно този етап може да се реализира като ***информационен модул***.

3.Моделиране. За да се управлява даден обект, той трябва първо да се моделира. Чрез модела информацията за неговите елементи и подeлементи се представя в удобна за възприемане форма. Предоставя се по-добър поглед върху същността на обекта на управление и се очертават някои от възможните алтернативи за неговото развитие и изменение. На този етап се разработва софтуерно приложение, което дава възможност да се конструира самият модел за управление. То трябва да бъде универсално приложимо и годено за моделиране на различни системи – икономически, социални, биомедицински и други.

Моделиращият модул се базира на схема на модел, която може да се използва при различни ситуации и етапи във времето, и която да се прилага за изучаване и управление на разнородни обекти. Тя представлява своеобразен макет за изработване на модели на системи, които не само ги описват в детайли, но са и лесни за разбиране и удобни за манипулиране от страна на управляващите.

4.Създаване на идеален модел. Определят се стойностите на параметрите, които характеризират желаното крайно състояние на системата.

5.Оценка на модела. На този етап се извъшва анализ на силните и слабите страни на модела, като се имат предвид позициите на предприятието, в което се създава. Съпоставят се текущото и желаното състояние на системата и се извеждат елементите и подeлементите, подлежащи на промяна, както и тези, които ще останат непроменени. Дава се отговор на въпросите с какви възможности раполага фирмата и има ли неразкрит или неизползван докрай потенциал.

Действията по точки 4 и 5 се включват в модула за моделиране.

6.Изработване на стратегия за действие. Изготвя се система от мероприятия, които следва да се предприемат, за да се реализира поставената за изпълнение цел. Тя се основава на анализ на модела чрез съпоставка на двете състояния (текущо и идеално), както и на редица аналитични механизми, даващи възможност да станат видими и да се оценят последиците от взетото управленско решение и да се откроят множеството възможни алтернативи.

Този етап покрива функцията планиране. Софтуерно може да се представи чрез ***модул планиране***, който включва помощни техники за изготвяне на бизнес план за дейността на предприятието като цяло, както и много други планове с различна степен на детайлизация по съдържащите се в тях действия.

7.Прилагане на приетата стратегия. Осъществява се стъпка по стъпка в зависимост от разработения план за действие по точка 6. Набира се екип от хора и същите се мотивират за работа. По време на този етап се проверява моделът в реални условия и се откриват допуснатите грешки, прави се оценка на постигнатите резултати и други. Покриват се функциите организиране, мотивиране и контролиране.

Необходимо е да се разработи модул ***организиране и контрол***. Той трябва да предоставя възможност да се следи за изпълнението на плана постъпково, да включва контролни точки за проверка и да сизгнализира при появата на грешки и отклонения извън допустимите норми, зададени в него.

Важен е също и ***модулът за мотивиране на персонала***. Базира се на система от мероприятия, под формата на санкции и награди, както и на персоналното стимулиране на личността. Съдържа механизми за оценка на работата на персонала и контролни точки, които помагат да се определи нивото на развитие на индивида, необходимостта от промяна на неговото възнаграждение или изпълняваната длъжност, от допълнително обучение и други. Наличието на този модул, заедно с модула за управление на конфликтите (който ще бъде разгледан по-долу), е предпоставка за ефективността на управлението на персонала, включително и в големите предприятия, където работят хиляди работници.

8.Отстраняване на допуснатите грешки. Извършват се корекции в модела и в стратегията за действие, ако това е необходимо. Покрива управленската функция регулиране. Преодолява се несъответствието между плановите и фактическите показатели.

На този етап се въдежда ***модулът регулиране***. Неговата цел е да предлага множество алтернативни решения за отстраняване на установените грешки, както и всички аргументи, които са довели системата до съответния избор.

9.Прилагане на обновената стратегия за действие. Покрива функциите организиране, мотивиране, контрол и регулиране.

10.Оценка на резултата и повтаряне на етапите от точки 1 до 10, докато изискванията бъдат напълно удовлетворени и се постигнат поставените цели. Осъществява се същинския преход към новото състояние на обекта на управлението.

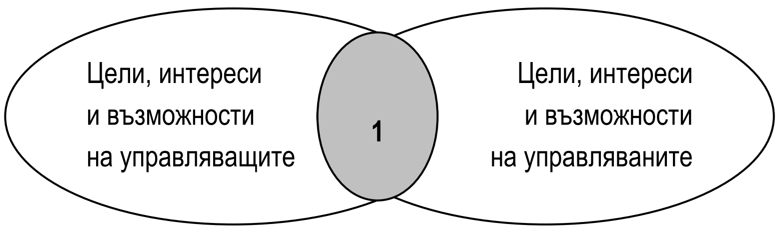
11. Стабилизиране на постигнатото състояние и вземане на решения за последващо развитие на системата. Прави се проверка във времето на това, което е постигнато. Преценява се дали е налице стабилна структура и доколко траен ще бъде достигнатия резултат. Ще устои ли той на натиска на средата и на факторите, които не са могли да бъдат взети под внимание при разработването на модела. Въз основа на това, както и на установените степени на свобода на обекта (области, върху които управляващият не може да окаже пълно или частично влияние), се очертават решенията за посоките в бъдещото развитието на обекта на управлението.

Управление на конфликтите

Освен модулите, които се включват в разгледаните по-горе етапи на управленския процес, е необходимо да се включи и още един – ***модул за управление на конфликтите***. Неговото предназначение е да бъдат определени и да се познават предварително възможните конфликтни точки и зони в предприятието и на тази основа да се преодоляват конфликтите още преди да са реално възникнали. Този модул трябва да притежава средства за разпознаване на предупредителни знаци за назряващи конфликти, както и инструментариум за тяхното недопускане. Може да се изгради и ***симулационен модул***, чрез който да се предвидят възможните ситуации при вземане на едно управленско решение (Reality Games). Така на принципа на игрите може да се извършва интерактивно обучение за управление на конфликтите.

Всяко предприятие се състои от хора, които го управляват и хора, които биват управлявани. Те постигат с общи усилия фирмената цел, работейки в хармония и допълвайки се взаимно. Между тях се осъществява непрекъснато взаимодействие, като подчинените също могат да оказват влияние върху управляващите.

Колкото в по-голяма степен съвпадат целите и интересите на управляваните и управляващите, толкова по-безконфликтно ще протича дейността на организацията. Противоречивите цели могат да породят сблъсък. Необходимо е да се определи зоната на съответсвията както на целите, така и на интересите и възможностите на двете страни в управленския процес (Виж фигура 1):



Фигура 1: Зона на съответствията

Ако фирмените цели се намират в Зона 1 (Виж фигура 1), тяхната реализация ще бъде максимално ефективна. Съгласуването на двете групи цели може да се постигне при осъществяване на редица психологически и финансови мероприятия, както и при адекватен подбор на персонала. Така посредством някои манипулативни механизми конфликтите могат да бъдат сведени до необходимия минимум. Потребността от последния е породена от обстоятелството, че ако те са малко на брой и предвидими, то биха могли да провокират постоянен стремеж към усъвършенстване. Това може да стане причина за растеж и фирмена самоорганизация. Важно е добре да се познават конфликтните зони.

Разработеното за тази цел софтуерно приложение следва да обхваща голям обем от информация, да я структурира, съпоставя и анализира. Като краен резултат то трябва да изведе потенциалните конфликтни зони, зоните на съответствията, както и да направи предложения за профилактика и преодоляване на потенциалните конфликти. Данните, които могат да се включат в модула, са следните:

- Индивидуалните особености на ръководителите от всички нива – лични интереси, мотивация за работа, успехи и провали, индивидуален опит, семейство, комуникативни умения и организаторски качества, лидерство, взаимоотношения, здравословно състояние и други, както и индивидуалните проблемни зони – кога и при какви условия е склонен към конфликти. Такава информация следва да се поддържа и за персонала, както и за предприятието като цяло.

- Целите на мениджърите, на персонала и общофирмените цели.
- Възможностите и опита на управляваните и управляващите. Това е техният личностен потенциал. От него зависи доверието, което подчинените имат към ръководителите, тяхната готовност и способност за изпълнение на поставените задачи и други. Важно е да се следи мениджърите да притежават повече възможности и по-висока квалификация от персонала. Обратното е предпоставка за конфликти поради неудовлетвореност от страна на работниците, които биха могли да създадат редица проблеми.
- Желание за личностно развитие и самоусъвършенстване на участниците в управленския процес. От това зависи бъдещото развитие на самото предприятие.

Изгражда се индивидуален модел на всеки от участниците в управленския процес. В процеса на работа тези модели се актуализират, обработват и съпоставят. Така става възможно да се изведат конфликтните зони и постоянно да се поддържа информация за актуалното състояние на обектите и субектите на управлението. Взетите по този начин управленски решения ще удовлетворяват както управляващите, така и управляваните и ще гарантират хармоничното функциониране на предприятието.

Изграждане на системата

От гледна точка на кибернетичната наука предприятието може да се дефинира като самоотганизираща се система, състояща се от управляваща и управлявана подсистема. Изхождайки от това становище, разширената информационна система за ръководителя следва да съдържа тези подсистеми и да копира техните функции и начин на работа, базирайки се на схематично представената структура на предприятието. Последната следва да отразява всички системни елементи, групирани по категории (Виж таблица 1).

Категории
Човешки ресурси
Дейности
Материална база
Структурни подразделения
Комуникационни канали
Софтуер
Други

Таблица 1: Основни категории на системата-предприятие

Въз основа на тези структурни единици могат да бъдат обособени отделни модули, които в последствие да се обединят и прерастнат в Extended EIS. По този начин първата категория от таблица 1 може да бъде изразена чрез Модул “Човешки ресурси”. Неговото съдържание е представено в таблица 2 (Виж таблица 2).

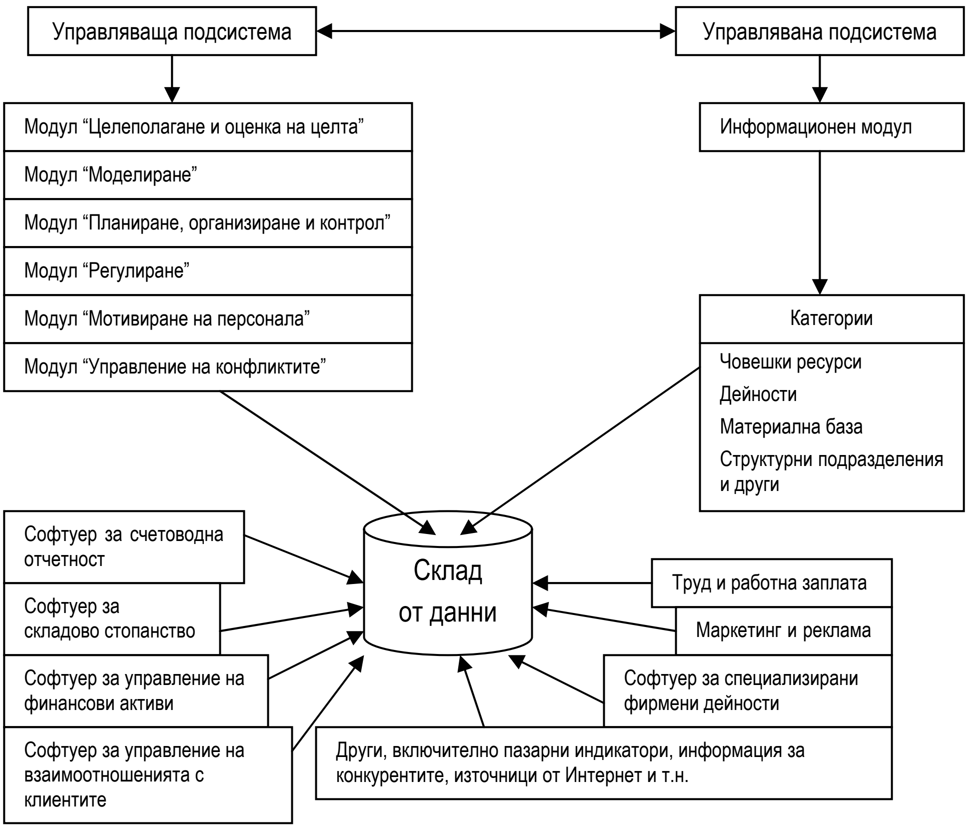
Модул “Човешки ресурси”	
Подкатегория	*Описание на подкатегория*
Персонални данни на лицето	Адрес, телефон и други.
Автобиографични данни	Образование, трудов стаж, интереси и други, преди постъпване на работа в предприятието.
Заемани длъжности в предприятието	Процес на израстване и успеваемост.
Мотивация	Причините, поради които работи в предприятието.

Семейно положение	Описание, степен на удовлетвореност и други.
Здравословно състояние	Преболедувани болести, хронични заболявания, предразположеност, общо физико-психическо състояние.
Лични качества	Посочват се такива като: комуникативност, умения за работа в екип, лидерски и организаторски умения, агресивност и други.
Потенциал	Използва ли докрай възможностите си? Как се справя в области, за които до този момент няма опит и информация? Притежава ли аналитичен ум? Желае ли да учи и да се развива? Как се справя и в кои области? и т.н.

Таблица 2: Съдържание на Модул "Човешки ресурси"

В предприятието следва да се създадат критерии за оценка на персонала, базирани на разработена система от показатели, които да се използват в Extended EIS. В последната намират отражение и данните от софтуера за заплати и други подобни, използвани в счетоводната отчетност.

Такова ниво на детайлизация се съдържа и във всички останали модули, чиято съвкупност представлява разширената информационна система за ръководителя. Последната използва информацията от наличните и в предприятието програмни приложения, като по този начин се избягва нейното дублирано въвеждане. Това може да стане най-добре при внедряване на технологията на складовете от данни, чието съществено предимство е, че обединяват данни от различни по характер източници. Използван като основа на Extended EIS, складът е средство, с помощта на което предприятието запазва прилагания до момента софтуер, надграждайки върху него новите технологии.



Фигура 2: Структура на разширената информационна система за ръководителя

В управляващата подсистема се съдържат основните бизнес правила, на които се подчинява цялата дейност на предприятието, както и средства за анализ и контрол. Тя следи за изпълнението на поставените цели, които се залагат в нея заедно с плановите показатели. Последните се сравняват с фактическите по каналите за обратна връзка от и към управляваната подсистема. Могат да се зададат степени на свобода на управляващата подсистема, които да и позволяват да прави промени в първоначално заложената фирмена цел въз основа на данните, предоставени от управляваната подсистема. Това би я превърнало в самоорганизираща се система със собствен интелект, аналогичен на този на мениджърите.

Заключение

Въз основа на всичко изложено до тук, може да се направи изводът, че разширената информационна система за ръководителя е способна до голяма степен да замени ролята на мениджърите, да облекчи тяхната дейност и да увеличи количествово свободно време – ресурс, който към настоящият етап на развитие на цивилизацията не достига на управляващите. Способства за повишаване на ефективността и качеството на фирменото управление и се характеризира с мултифункционалност и мултидисциплинарност. Това я превръща в информационна система на бъдещето, базирана на софтуерната реализация на мисловната дейност на хората.

Библиография

[1] Христов С., Основи на управлението, ИУ – Варна, 1997, стр.7

[2] Олдер Г., Хэзер Б., NLP – полное практическое руководство, София, 2000, стр.37

[3] Моцев М., Информационни системи в бизнеса, Университетско издателство “Стопанство”, София, 1994, стр.44

Информация за автора

Тодорка Ковачева - Икономически Университет – Варна, докторант към Катедра “Информатика”; e-mail: todorka_kovacheva@yahoo.com

КООРДИНАТНО БАЗИРАНА ФИЗИЧЕСКА ОРГАНИЗАЦИЯ ЗА КОМПЮТЪРНО ПРЕДСТАВЯНЕ НА ИНФОРМАЦИОННИ ПРОСТРАНСТВА

Красимир Марков

***Резюме**: В статията се представя най-новата реализация на един метод за координатно организиран достъп до свръхголеми информационни пространства.*

***Ключови думи**: Многообластен информационен модел и метод за достъп, Координатно базирани информационни пространства.*

Многообразие и подредба

Светът, който ни заобикаля може да се характеризира с една дума като «многообразие».

Многообразие, съществуващо тук и сега.

Без минало и без бъдеще.

Странно за нашето възприемане на този свят като четиримерно съществуване.

Странно, защото нашата памет и нашето съзнание градят съвсем друга картина - на пространственост и протяжност.

Едва ли бихме се съгласили, че в света няма така нужната ни подредба.

Подредба, създавана през хилядолетията, поддържана и развивана постоянно.

Ако настоящият текст беше художествено произведение, то може би този стил на изложение би бил интересен и интригуващ. Но в научния текст следваме друга «последователност» и «подредба». Безспорно – условна, защото така сме се условили при изграждането на нашата ***комуникативна конвенция***, срещу разрушаването на която бихме се противили.

Но ако става дума за нейното развитие и усъвършенствуване?

Нека не сме големи оптимисти...

Подредеността на света е информационен феномен. Ние подреждаме света, защото можем да подредим ***информационните модели*** за него. Нещо повече – ние действуваме следвайки нашето ***информационно очакване***, породено на база построените до момента информационни модели [Markov et al, 2003].

И градим света около нас следвайки нашата информационна представа за това какъв бихме искали да е.

Подреден ...!

Информационното взаимодействие

Понятието «информационно взаимодействие», както и съпътствуващите го понятия «информационен обект», «информационен контакт» и др. бяха въведени преди повече от двадесет години в [Markov, 1984] и днес те са част от ежедневно употребяваната терминология (вж. например [Бърнев и др., 1988, стр.9-10]). Отделен раздел на "Обща теория на информацията" [Markov et al, 2003] е посветен на именно на тези понятия. Ще припомним накратко някои основни постановки.

Ние непрекъснато градим ***информационни модели*** за света и за себе си в този свят. Многообразието на информационни модели съответствува на многообразието на света, но е значително по-малко мощно. От друга страна, координирането на действията ни с други хора изисква постоянна информационна обмяна, в основата на която са ***информационните обекти***. Това са информационни модели, реализирани по определен начин в окръжаващата среда, които биха могли да се възприемат от другите.

Ще припомним общата схема на информационното взаимодействие.

Един човек (или група хора) S_1 реализира в околната среда конкретен информационен обект O_i, съответстващ на даден информационен модел, който е породен в неговото съзнание, и който би могъл да се възприеме от друг човек (или група хора) S_2 :

$$S_1 \rightarrow O_i \rightarrow S_2 \quad (1)$$

Реализацията O_i е ***информационен обект*** (**ИО**), само ако съществува ***конвенция*** между източника и получателя, на базата на която възприетия информационен модел би се съотнесъл с вече наличните в съзнанието на възприемащия. В противен случай тя би била само едно въздействие върху околната среда, което би предизвикало временни или постоянни изменения на обекти от нея.

Релациите (1) определят един ***информационен контакт***. С други думи, информационен контакт е налице само ако съществува съответна информационна конвенция между S_1 и S_2 , и тройката

$$(S_1, O_i, S_2)$$

е свързана с релациите (1).

Съвкупността от информационните контакти между S_1 и S_2, и, респективно, между S_2 и S_1 формират ***информационното взаимодействие*** между тях.

Съвкупността от всички ИО, чрез които се осъществява информационното взаимодействие формира неговата ***информационна база***.

Погрешно е да се счита, че информационните обекти са феномен само при хората. Но само при хората, например, съществуват писменост и, съответно, ***текстови информационни обекти***.

Многообразието на света предизвиква многообразие на информационните модели, респективно на тяхните реализации като информационни обекти. Ние възприемаме реалния свят като пространствена йерархия, развиваща се във времето, и следователно информационните модели и съответстващите им ИО също са свързани в ***динамична пространствена йерархия.***

Многомерност

Много важна особеност на информационните обекти е наличието на ***специфични характеристики (атрибути)***. Характеристиките се определят от хората, участващи в процеса на информационно взаимодействие. Ако всички атрибути на два информационни обекта имат едни и същи стойности, то това ги прави неотличими един от друг.

Ако множеството от стойности на даден атрибут бъде ***подредено и номерирано***, то този атрибут се превръща в «***размерност***», чрез която ИО, които му съответстват също могат да бъдат подреждани.

Ако един ИО се характеризира от единствен атрибут той се нарича "***едномерен***". Естествено, обектът се нарича "***многомерен***" ако се характеризира от множество атрибути. Многомерният ИО, определен от k атрибути може да бъде представен като "точка" в k-мерно пространство. Това ореление напълно се съгласува с виждането в [Sitarama et al, 1988].

Така идваме до простата идея, че информационната база би могла да се организира въз основа на една или повече размерности и по този начин да се разглежда като съответно-мерно ***информационно пространство***. Очевидно, многомерното информационно пространство ще се характеризира от повече от една размерност.

В простите случаи «мерността» на информационната база ще е ***обозримо малка***, например, в случай, че описваме литературата към настоящия текст.

Но ако трябва да организираме информационна база за взаимодействията в реална, например – бойна, ситуация, в която са ангажирани хиляди елементи – хора, техника, поземни и надземни съоръжения, собствени и на противника, който в наши дни е навсякъде, а не просто зад една въображаема фронтова линия, и при което всеки елемент се характеризира от голямо количество индивидуални атрибути, а информацията непрекъснато постъпва както от човешкия ресурс, така и от съответни датчици на техническите системи и съоръжения в цялото пространство...?

Трудна, но интересна и значима задача.

Решението, което като че ли само идва, е

на многообразието да се отговори с многообразие!

И с ред...!

> ***Многообразие от информационни обекти, които да се организират в практически неограничено, динамично, йерархично подредено информационно пространство.***

Реализуемо ли е това?

Като че ли природата е отговорила положително чрез нас самите, много преди да има компютри и когато **MICROSOFT®** «го нямаше даже».

Е, като се има пред вид обема на възможната дискова памет и капацитета на файловите системи на компютрите, с които разполагаме, думите «практически неограничено» придобиват съвсем ограничен смисъл! Така, че задачата ни трябва да се произнесе може би като се замени «практически неограничено» с «ефективно» по отношение на памет и време в условията на конкретна компютърна структура.

В добавка – нека отчетем, че много от потенциалните ИО просто не съществуват и нашето ***информационно хиперпространство*** реално е ***почти празно***. При това, «не съществува» не означава, че в съответната хиперточка има определено количество интервали (празни символи), а просто няма нищо и за тази хиперточка не се заделя памет, а времето за достъп до "нищото" е пренебрежимо малко в сравнение с времето за достъп до реално съществуващ информационен обект.

Накрая, що се касае до реализация на компютри от типа IBM PC, нашето пространство ще трябва да може да се разположи в рамките на ***един файл*** (в термините на MS DOS, MS WINDOWS и LINUX), тъй като апарата за достъп до файловете и съдържащите ги структури, удобен за човека, е ***крайно неефективен*** при автоматизирана работа с подредени информационни пространства, съдържащи милиони ИО.

В периода от 1974 г. до 2004 г., с промяната и развитието на компютърната техника, задачата постепенно се разширяваше и обхващаше нови елементи, които усилваха първоначалния проект.

Тук ще я поставим в нейния вариант, който съответствува на състоянието към 2004 год. и ще обсъдим едно възможно решение, в чиято основа стои единен подход за физическа организация, базиран на разбирането, че информацията трябва да се съхранява ***подредено*** и да се достига чрез ***числови координати***. Този модел на организация беше наречен «**Многообластен информационен модел**» (**МИМ**) [Markov 2004], а съответното му програмно осигуряване - «**Многообластен метод за достъп**» (**ММД**) [Markov 1984].

Постановка на задачата

Нека n е размерността на регистрите на процесора на компютъра (напр. за 32-битовите компютри: n=32).

Да се проектира и реализира метод за достъп, който да позволи в рамките на един архив (един файл или един раздел (disk partition) за операционните системи MS DOS, MS WINDOWS и LINUX) да се съхранят произволно количество информационни елементи, подредени по и достъпни чрез цялочислени координати, чийто брой е фиксиран за конкретен елемент, но е променлив в рамките на архива.

Елементите в един архив са последователности от байтове с променлива дължина от 0 (нула) до 2^{n-2} (за IBM PC : 2^{30} = 1 073 741 824, т.е 1G). Няма ограничение за броя на елементите в един архив, но сумарната им дължина плюс дължината на служебната информация за индекси и пр. не могат да надхвърлят 2^n байта (4G байта за файловите системи за момента).

Елементите от един архив следва да се организират в архивно информационно пространство на базата на динамична структура от координатни системи. Координатните системи могат да са едномерни, двумерни и т.н. Няма особено ограничение за броя на размерностите на координатните системи – те могат да са с размерност от 0 до 2^n, но за момента и 100 е напълно достатъчно.

В един архив трябва да е възможно да има повече от един вид координатни системи, но един конкретен елемент да може да се достигне само чрез една координатна структура. Това означава да се работи на принципа на позиционните бройни системи, за които по подразбиране старшите нули се пропускат. Позиционната система, на чиято основа да се изгражда реализацията следва да е с основа 2^n. Така всеки елемент ще се достига чрез определен брой координати, които са n битови цели числа.

Координатите да се задават чрез масив**.** Броят на координатите винаги да се указва на нулево място в масива, след което се задават съответните координати. Мястото на стойностите на координатите в масива има значение на позиция в позиционна система с основа 2^n.

Така, всеки елемент ще се достига чрез координатна m-торка, $m=0,\ldots,2^n$, но по същество той се разполага в една точка от координатната ос, която се описва чрез цели числа от позиционна бройна система с основа 2^n.

Предполага се, че архивното пространство няма да бъде запълнено напълно. Нещо повече, огромната координатна мощ ще се използва само в съответствие с конкретните потребителски нужди и, като правило архивното пространство ще е почти празно.

Така от една страна имаме неограничена възможност да адресираме в архивното пространство, и ограничен размер на архивите, от друга.

Пространството следва да е организирано йерархично: празното подпространство е елемент на пространството, а елементите се агрегират в подпространства по нарастване старшинството на координатите и нарастване на размерността.

ArM32®

Решението на поставената задача е предложено в [Markov, 2004]. През последните двадесет години то стана популярно по името “Многообластен информационен модел” [Markov, 1984], [Markov, 1985]. Настоящата статия представя най-новата програмна реализация на модела, известна под името **ArM32®**

ArM32® се изгражда по единна информационна технология, осигуряваща достъп чрез цифрови координати до полупълно динамично информационно пространство.

Архиви

Всеки архив е един файл с блокова структура с пряк достъп. Размерът на блока в един файл е **512B**.

Организация на архив:

1). Файл с постепенно нарастваща дължина по 1 блок.

Операции:

- запис на блок (в рамките на файла / в края на файла);
- четене на блок.

2). Върху множеството от блокове с пряк достъп се изгражда следната структура от данни:

- множество от всички заети блокове във файла;
- множество от всички свободни блокове във файла.

Операции:

- вземане на свободен блок;
- освобождаване на блок.

Няма подразбиране за наставката към името на архив.

Предполага се, че всички файлове, които са архиви, са отворени постоянно. (Това ще рече, че са винаги достъпни и всяка нова информация се отразява върху диска веднага).

Елементи

Елементите в един архив са последователности от байтове с променлива дължина от 0 (нула) до 2^{30} (1 073 741 824), т.е 1G.

Няма ограничение за броя на елементите в един архив, но сумарната им дължина плюс дължината на служебната информация за индекси и пр. не могат да надхвърлят 4G байта.

Архивно пространство и координати

Елементите от един архив се организират в архивно информационно пространство на базата на координатни системи. Координатните системи могат да са едномерни, двумерни и т.н. Няма ограничение за броя на размерностите на координатните системи (но и до 100 засега е достатъчно).

В един архив е възможно да има повече от един вид координатни системи, но един конкретен елемент може да се достигне само чрез една координатна структура. Това се постига на принципа на позиционните бройни системи, за които по подразбиране старшите нули се пропускат.

Позиционната система, на чиято основа се изгражда ArM32 е с основа 2^{32}. Така всеки елемент се достига чрез определен брой координати, които са 32 битови цели числа (cardinal).

Координатите се задават чрез масив от тип **cardinal.** Броят на координатите винаги се указва на нулево място в масива, след което се задават съответните координати. Мястото на стойностите на координатите в масива има значение на позиция в позиционна система с основа 2^{32}.

Така, всеки елемент се достига чрез координатна n-торка, но по същество той се разполага в една точка от координатната ос, която се описва чрез цели числа от позиционна бройна система с основа 2^{32}.

ArM индекси

Предполага се, че архивното пространство няма да бъде запълнено напълно. Нещо повече, огромната координатна мощ ще се използва само в съответствие с конкретните потребителски нужди и като правило архивното пространство ще е почти празно.

Така от една страна имаме неограничена възможност да адресираме в архивното пространство, и ограничен размер на архивите (4G), които са файлове. Това налага създаване на структура на съответствие между координатите на елементите и тяхното реално място във файла. Тази структура се нарича **"ArM-индекси".** В текста по-долу ако не се уточнява друго понятието "индекс" означава точно "ArM-индекс".

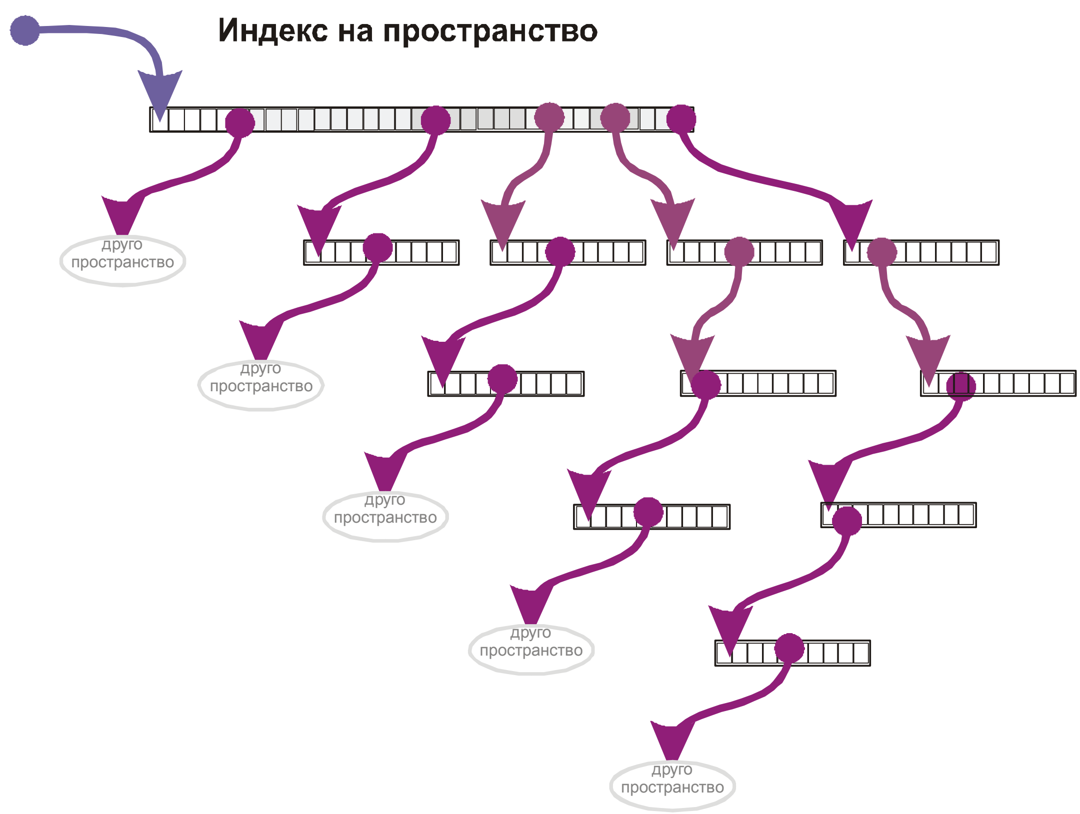
Индексите за една координата се изграждат чрез въвеждане на ограничения и допълнения в дървовидна структура от блокове с указатели максимум до пет нива (фиг.1).

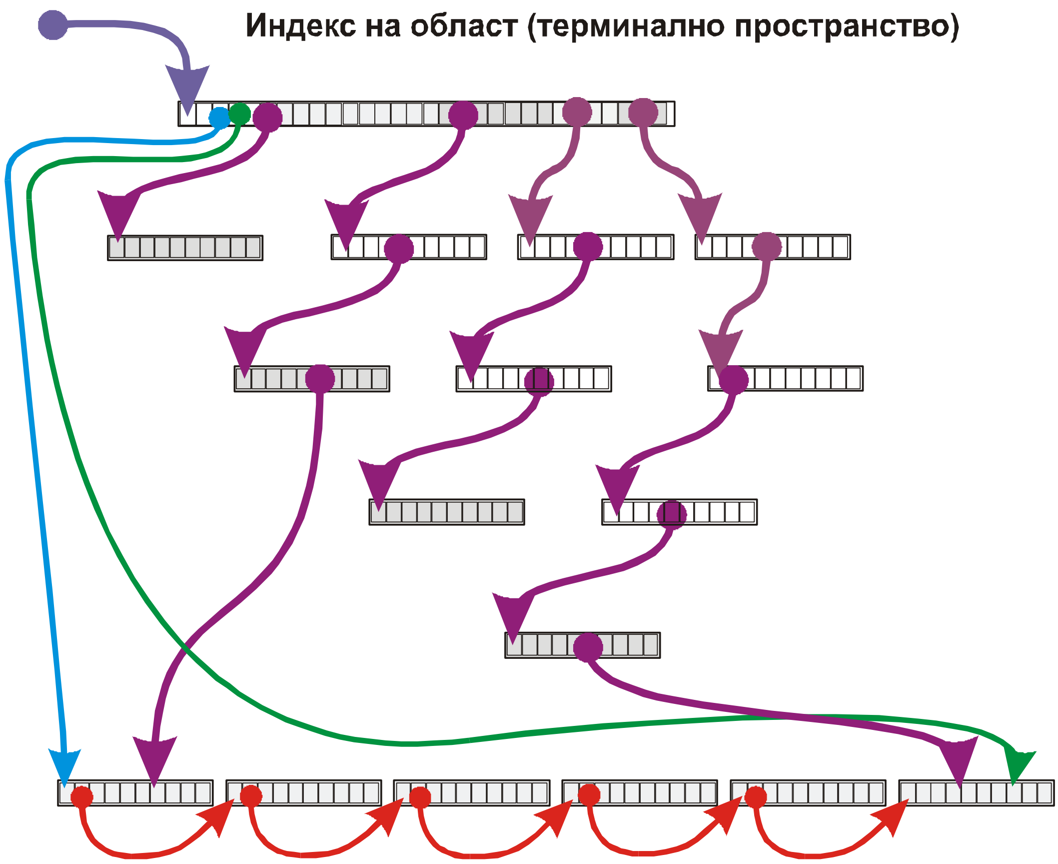
Всеки индекс (едно измерение) се нарича **"пространство"**. Терминалните пространства (индекси към елементи, а не към други индекси) се наричат едномерни пространства или по-просто **"области"**.

Така, физическата организация на **ArM32** се строи на следните три нива:

- пространства;
- области;
- елементи.

По-долу са представени двете основни индексни дървовидни структури на **ArM32**.


Индекс на пространство
друго пространство
друго пространство
друго пространство
друго пространство
друго пространство


Индекс на област (терминално пространство)

Типове блокове в ArM32

Всеки архив е един файл с блокова структура с пряк достъп. Размерът на блока в един файл е **512B**. Основните типове блокове са представени в следващата таблица.

тип	наименование	Определение
0	**Главен блок** TZeroBlock	Нулевият блок на файла съдържа служебна информация на архива и потребителя. Сочи към свободното пространство от блокове в архива. Съдържа семафори и пароли за синхронизация при съвместна работа на няколко процеса.
1	**Глава на архива** TSpaceBlock	Съдържа пет зони от преки указатели към глави на пространства или междинни индексни блокове (по структура съвпада с тип 2).
2	**Глава на индекс на пространство** TSpaceBlock	Съдържа пет зони от указатели към други пространства: - преки указатели - нулева зона - непреки указатели през един, два, три или четири междинни индексни блока (зони от 1 до 4).
3	**Междинен индексен блок за пространство** TIndexBlock	Съдържа 169 трибайтови указатели към блокове от архива.
4	**Глава на индекс на терминално пространство (област)** TDomainBlock	Съдържа: - информация за списъка с данни; - четири зони от непреки указатели през нула, един, два или три междинни индексни блока (от тип 5), водещи до блокове с указатели към елементи (от тип 6).
5	**Междинен индексен блок за област** TIndexBlock	Съдържа 169 трибайтови указатели към блокове от архива.
6	**Блок с указатели към елементи** TelementBlock	Съдържа 126 четирибайтови указатели към елементи (номер блок + логическо отместване)
7	**Блок с данни** TDataBlock	Съдържа 503 байта с данни от елементите на областта. Всеки елемент се състои от дължина (1 до 4 байта) и стойност от байтове с размер от 1B до 1GB, указан от дължината. Първите два бита в дължината определят нейната дължина в байтове (00h : 1 байт, 01h : 2 байта, 10h : 3 байта, 11h : 4байта).

Структура на свободната памет на архив

По време на работа с архив е възможно да се освобождават блокове, които не са необходими за съхраняване на информация. Тези блокове не се връщат на операционната система, а се включват в структурата на свободната памет в архива. Тази структура се обслужва от указатели от главния блок на архива:

- Указател към началото на свободната памет;
- Указател към блок за взимане от свободната памет;

При освобождаване (изтриване) на дадена структура от архива (индекс на пространство или област, междинен индексен блок, блок с указатели към елементи) тя се премества в началото на структурата на свободната памет. За целта указателят към началото на свободната памет и досегашната първа освободена структура се насочват към главата на освобождаваната структура. Така тя застава на първо място.

По този начин свободната памет се изгражда като едносвързан списък от глави на различни структури. Контролът, който в другите структури се базира на двусвързаност, тук се осъществява чрез евристиката, че всяка глава сочи към нулев байт от блок на друга глава или към 160H на нулевия (главния) блок на архива.

Координатен масив

Всички логически операции се обслужват от един или повече координатни масиви и други параметри - буфери, дължини, отмествания, нива и др.

Координатният масив е наредена N-орка от числа тип **cardinal**.

Нулевият елемент от масива указва колко индексни прехода ще трябва да се извършат докато се достигне до стойността на елемента, който се указва чрез този масив. Този елемент се нарича "размерност на пространството", в което ще се работи. Няма ограничение за размерността на пространството - тя може да е от 1 до 2^{32} −(като в първия блок се поместват указатели към първите 100 размерности).

Примерен координатен масив е представен на таблицата по-долу:

N	Размерност на пространството
k_n	Най-старша координата
k_{n-1}	…
…	
K_1	Номер на елемента

Забранено е да се указват нулеви стойности на координатите, с изключение на операциите за работа с подпространства, където координатите след определено ниво се игнорират (нямат значение).

Основни логически операции на ArM32

Управление на процесите

- Проверява дали е има връзка с **ArM32.DLL** ArmInstalled
- Връща кода на грешка ArmStatus
- Връща сигнатурата на **Arm32** ArmVersion

Управление на архивите

- Управление на достъпа до архив ArmLock
- Информация за архив ArmInfo
- Профилактика и възстановяване на архив ArmRepair

Координатни операции на ниво елемент

Основните операции за достъп до елементите на архива ползват координатен масив за указване мястото на елемента в информационното пространство. Този универсален начин за достъп отнема времеви ресурси за откриване на физическото място на елемента. Това е оправдано при предварително неопределен стил на достъп до елементите.

Основните координатни операции са:

- Получаване дължина на елемент ArmLength
- Четене на елемент ArmRead
- Запис на цял елемент ArmWrite
- Добавяне към елемент ArmAppend
- Вмъкване в елемент ArmInsert
- Премахване на част от елемент ArmCut

– Заместване част от елемент ArmReplace
– Изтриване елемент ArmDelete

Дескрипторни операции в ArM32

При многократно обръщение към един и същ елемент не е оправдано да се извършва всеки път локализация чрез координатен масив. Така се достига до идеята да се използва специален дескриптор, в който да се запише информация по физическото разположение на елемента в архива и чрез него да става достъпа до елемента. Операциите на ArM32, които ползват дескриптор за достъпа до елемент се наричат "дескрипторни операции". Основните дескрипторни операции са:
– Позициониране в елемент ArmSeek
– Пряко четене от елемент ArmGet
– Пряко писане в елемент ArmPut

Операции на ниво пространство (в индекси)

– Следващ Наличен ArmNextPresent
– Следващ Празен ArmNextEmpty
– Предишен Наличен ArmPrevPresent
– Предишен Празен ArmPrevEmpty
– Брой компоненти ArmCountSpace
– Изтриване подпространство ArmDelSpace

Операции, при които могат да участват и два архива

– Копиране подпространство ArmCopyClear, ArmCopyMerge
– Преместване подпространство ArmMoveClear; ArmMoveMerge

Заключение

В статията се представя най-новата реализация на един метод за координатно организиран достъп до свръхголеми информационни пространства. До този момент реализациите преминаха през десет версии, които е възприето да се номерират от нула:

No.	години	машина	Език за програмиране	Наименование
0	1974-75	МИНСК 32	Assembler	**MDAM 0**
1	1980-81	ЕС 1040	FORTRAN	**MDAM 1**
2	1982-83	СМ4	FORTRAN:	**MDAM 2**
3	1984-85	СМ4	Assembler:	**MDAM 3**
4	1985	Apple II	UCSD Pascal	**MDAM 4**
5	1986-1990	IBM PC	Assembler с интерфейс към Pascal, C, Lisp	**MDAM 5**
6	1988	SUN	C	**MDAM 6**
7	1990-1993	IBM PC	Assembler:	**ArM 16 DOS**
8	1998	IBM PC	DELPHI за MS Windows 95	**ArM 16 WIN'95**
9	2003	IBM PC	DELPHI за MS Windows XP	**ArM 32**

Реализациите с имена **MDAM** и **ArM 16 WIN'95** имаха експериментален характер. Реализациите **ArM 16 DOS и ArM 32** са промишлени, като до момента **ArM 16** има над хиляда инсталации в България. **ArM 32** се подготвя да замени **ArM 16**.

Благодарности

В реализациите на метода през годините се включиха:

- Димитър Гелев (MDAM 4)
- Тодор Тодоров (MDAM 5 с интерфейси за PASCAL и C)
- Васил Николов (реализация на интерфейс към MDAM5 за LISP)
- Васил Василев (ArM 16 DOS и ArM WIN'95)
- Илия Митов и Красимира Иванова (ArM 32).

Считам се приятно задължен да им изкажа своята благодарност за плодотворното сътрудничество.

Версиите MDAM са разработвани като научни задачи в Института по математика и информатика при БАН. Реализацията на версиите ArM са финансирани от и са собственост на фирмите ФОИ-КОМЕРС, ФОИ КРЕАТИВ и ФОИ КОНСУЛТ.

Литература

[Markov 1984] Kr.Markov. A Multi-domain Access Method. // Proceedings of the International Conference on Computer Based Scientific Research. Plovdiv, 1984. pp. 558-563.

[Markov, 1985] Kr.Markov. A Microsystem for information Contact. // Mathematics and Education in Mathematics, BAS, SOFIA, 1985, pp.471-475.

[Markov et al 2003] Kr.Markov, Kr.Ivanova, I.Mitov. General Information Theory. Basic Formulations. // FOI-COMMERCE, Sofia, 2003.

[Markov, 2004] Multi-Domain Information Model. Proceedings of XXIX International Conference "Information and Communication Technologies & Programming" (ICT&P 2004). Sofia,FOI-COMMERCE, 2004, pp. 79-88.

Информация за автора

Красимир Марков: Институт по математика и информатика - БАН;

п.к. 775, Sofia-1090, Bulgaria; e-mail: foi@nlcv.net

ЕДИН ПОДХОД ПРИ АНАЛИЗ НА ПРЕДПОСТАВКИ ЗА ИЗГРАЖДАНЕ НА Е-ОБЩИНА

Димитър Николов, Румяна Стефанова, Ивелин Стефанов

А. Идейна предпоставка

I. МОТИВ: Основният мотив е, че електронното правителство не е ЦЕЛ, а средство за постигане на отворено обществено управление с естествената прозрачност, технологично обусловената ефикасност и интегритета си.

II. ИДЕЙНИ ЦЕЛИ В РЕАЛИЗАЦИЯТА НА eGovernment:

1. Политическа и интеграционна цел – за създаване на обществено управление, което да се интегрира в европейските структури.
2. Прозрачност и проследимост на решенията, гарантираща публичност на вземане на решенията и подкрепа на масмедиите.
3. Търсене и намиране на услугите. Персонализиране на обществената услуга за всеки, за да се повишава търсенето, устойчивото развитие и надеждността на реализацията й. Пазарният характер на тълкуването на тази идея противоречи на широката обхватност на тази услуга – за всеки слой, социална и етническа група.
4. Организационният ефект върху увеличаване на вътрешната ефективност, вследствие на организираността на обслужването и създаване на нови идеи и решения за обслужване
5. Икономически ефект от повишената продуктивност и резултатност, включително и на събираемостта на данъци и др.
6. Социалният ефект на равнопоставеност пред информацията, защото клиентите на общинското обслужване не могат да се избират, те са гражданите ни.

Това е общото идеологическо пространство на проблематиката.

В. Необходимост

III. ПРАКТИЧЕСКИ НАСОКИ: Практическите насоки са дадени в стратегията за реализиране на електронно правителство. В приложение 3 са дадени конкретните индикативни 20 услуги, изпълнението на които ще гарантират пътя към реализацията на тази стратегия.

Индикативни административни услуги

Като индикатори за оценка на развитието на електронното правителство Европейската комисия прие следния списък от on-line изпълнявани основни административни услуги – 12 за граждани и 8 за бизнеса:

Административни услуги за граждани:

1. Подоходни данъци: декларации, уведомяване.
2. Услуги по търсене на работа при бюрата на труда.
3. Социални осигуровки, помощи за безработни, добавки за деца, медицински разходи, стипендии.
4. Лични документи (паспорти, свидетелства за управление на МПС).
5. Регистрации на МПС (нови, използвани, внесени МПС).
6. Подаване на документи за строителни разрешения.
7. Декларации към полицията (напр. при кражба).
8. Публични библиотеки (каталози, машини за търсене).
9. Свидетелства (за раждане, брачни и др.).
10. Дипломи за средно и висше образование.
11. Смяна на адресна регистрация.
12. Услуги, свързани със здравеопазването (интерактивни съвети относно наличността на определен тип услуги в различните болници; запазване на час за преглед).

Административни услуги за бизнеса:

1. Социални осигуровки за заетите.
2. Корпоративни данъци: декларации, уведомяване.
3. Данък върху добавената стойност: декларации, уведомяване.
4. Регистрация на нова фирма.
5. Изпращане на данни до Националния статистически институт.
6. Митнически декларации.
7. Разрешения, свързани с екологични изисквания (включително докладване).
8. Обществени поръчки.

Прогресът в предоставянето на тези услуги ще се оценява в следните 4 степени:

1. **Предоставяне на on-line информация за услугата.**
2. **Еднопосочно взаимодействие.**
3. **Двупосочно взаимодействие.**
4. **Цялостно обслужване on-line, включително доставка и разплащане.**

IV. ЗАКОНОВИ ЗАДЪЛЖЕНИЯ НА ОБЩИНИТЕ, ИЗВЪН ИНДИКАТИВНИТЕ УСЛУГИ- ИЗИСКВАНИЯ НА ЗАКОНА ЗА ЗАЩИТА НА КЛАСИФИЦИРАНАТА ИНФОРМАЦИЯ

Конституцията от 1991г. очертава България като парламентарна република с местно самоуправление. Местната власт е равнопоставен партньор на централната държавна власт. Извършващите се демократични промени в нашето общество отреждат все по-активно участие на органите на местното самоуправление в обществения живот. Чрез провежданата децентрализация държавата предоставя нови дейности на общините. Една от най-важните задачи за общините е да осигурят на гражданите качествени и леснодостъпни административни услуги. Услугите са свързани с предоставяне на информация, засягаща законни права и интереси на гражданите. Редът и условията за предоставяне на съответната информация, се уреждат със закона. През последните години при предоставяне на услуги от общините се внедряват съвременни автоматизирани информационни технологии. Бъдещото присъединяване на страната ни към Европейския съюз и новите европейски изисквания към предоставянето на услуги и информация на гражданите и обществото, наложиха приемането на нови закони. Все повече се повишават изискванията на гражданите към общините за получаване на качествени и навременни административни услуги.

Законова рамка

- *Конституция на Република България* /обнародвана "ДВ" брой 56 от 13.07.1991г/. Основен закон на държавата. Конституцията прогласява, че всички граждани на Р България са равни пред закона. Не се допускат никакви ограничения на правата или привилегии, основани на раса, народност, етническа принадлежност, пол, произход, религия, образование, убеждения, политическа принадлежност, лично и обществено положение или имуществено състояние. Всеки има право да търси, разпространява или получава информация. Осъществяването на това право не може да бъде насочено срещу правата и доброто име на другите граждани, както и срещу националната сигурност, обществения ред, народното здраве и морала. Гражданите имат право на информация от държавен орган или учреждение по въпроси, които представляват за тях законен интерес, ако информацията не е държавна или друга защитена от закона тайна или не засяга чужди права.
- *Закон за местното самоуправление и местната администрация* /обнародван "ДВ" брой 77 от 17.09.1991г, многократно изменян и допълван/. Урежда обществените отношения, свързани с местното самоуправление и местната администрация. Орган на изпълнителната власт на общината е кметът, който се избира пряко от населението. Кметът организира изпълнението на задачите, които произтичат от законите. Кметът утвърждава Устройствен правилник на общинската администрация, в който конкретно регламентира и функциите и задачите на дирекциите, отделите и секторите, които работят по предоставяне на административни услуги на гражданите. Кметът назначава безсрочно секретар на общината, който организира дейността на общинската администрация, деловодството, документооборота и общинския архив. Кметът на общината в случаите, определени от закона, изпълнява и функции, възложени му от централните държавни органи.

- *Закон за административно обслужване на физическите и юридическите лица* / обнародван "ДВ" брой 95 от 02.11.1999г/. Законът урежда реда за административното обслужване на физически и юридически лица, както и обжалването на отказа за предоставяне на услуга. Въвежда легално понятие "административна услуга". Задължава органите на изпълнителната власт /държавна и местна/ да предоставят административни услуги, при поискване от заинтересовано лице, когато услугата е от значение за негови законни права и интереси. Отказът да се извърши поискана услуга се извършва с мотивиран акт, който подлежи на обжалване.
- *Закон за гражданска регистрация* /обнародван "ДВ" брой 67 от 27.07.1999г, многократно изменян и допълван/. Законът урежда ред и условия за гражданската регистрирация на физическите лица, чрез вписване на събитията раждане, брак и смърт в регистрите на населението. Регламентира данните в актовете за гражданското състояние. Въвежда Единна система за гражданска регистрация и административно обслужване на населението /ЕСГРАОН/, автоматизирани информационни фондове на ЕСГРАОН, ред за предоставяне на данни от тази система, както и тяхната защита. Правата и задълженията на органите на гражданската регистрация на населението, в т.ч. и на общините.
- *Закон за електронния документ и електронния подпис* /обнародван "ДВ" брой 34 от 06.04.2001г, в сила от 07.10.2001 г/. Законът урежда легални понятия "електронен документ" и "електронен подпис", ред и условия за предоставяне на удостоверителни услуги по електронен път. Създава защита на тайната на данните за създаване на електронен подпис. В глава V се регламентира приложението на електронния документ и на универсалния електронен подпис от държавата и от общините. Съгласно закона, приемането и издаването на електронен документ, подписан с универсален електронен подпис от органите на местното самоуправление, се уреждат с техен акт. Редът и формата за извършване и съхраняване на електронни документи се урежда с вътрешни правила. Защитата на личните данни от доставчиците на удостоверителни услуги и защитата на водените регистри се урежда с отделен закон.
- *Закон за авторското право и сродните му права* /обнародван "ДВ" брой 56 от 29.06.1993г, многократно изменян и допълван/. В Раздел VII "Използване на компютърни програми" е въведен нов чл. 71а /"ДВ" брой 77 от 2002 г/. Регламентирано е, че управлението и контролът върху използването на придобитите софтуерни активи от органите на държавната власт и техните администрации, както и от органите на местното самоуправление и техните администрации се извършва по ред, определен от Министерския съвет.
- *Закон за достъп до обществена информация* /обнародван "ДВ" брой 55 от 07.07.2000г/. Законът урежда обществените отношения, свързани с правото на достъп до обществената информация и въвежда легално понятие за такава информация. Задължени субекти по този закон да осигуряват достъп до обществена информация са държавните органи и органите на местното самоуправление. Всеки гражданин на РБългария има право на достъп до обществена информация по ред и условия, уредени в този закон, освен ако с друг закон не е предвиден специален ред за търсене, получаване и разпространение на такава информация. Уреден е ред за обжалване на отказите за предоставяне на достъп.
- *Закон за защита на класифицираната информация* /обнародван "ДВ" брой 45 от 30.04.2002г/. Законът урежда обществените отношения, свързани със създаването, обработването и съхраняването на класифицираната информация, както и реда и условията за получаване на достъп до тази информация. Създаден е централен държавен орган за защита на класифицираната информация: Държавна комисия по сигурността на информацията. Регламентирани са функциите на службите по сигурност във връзка със защитата на класифицираната информация. В изпълнение на този закон, кметовете на общини са задължени да назначат на свое подчинение служител по сигурността на информацията. Законът урежда видовете класифицирана информация: държавна тайна, служебна тайна, чуждестранна класифицирана информация. Регламентирани са и нивата на информация и техния гриф за сигурност : "строго секретно", " секретно", " поверително", "за служебно ползване".

C. Целта

За да се намерят пътищата на оценката за желанията, приоритетност на инициативите според необходимостта, изискванията и възможностите за поставяне на една задача за електронна община трябва да се изследват НАГЛАСИТЕ И ГОТОВНОСТТА в ОРГАНИЗАЦИЯТА И РАБОТНАТА ПРАКТИКА НА ОБЩИНАТА в административното и обкръжение и ред на вътрешна организация.

Целта на това изследване е да се извлече яснота какви приоритети намира общината за значими за себе си, какви цели си поставя, какви възможности има за реализирането им. Резултатът е да се намери пространство от възможности и ограничения, в което да се проектира икономическата целесъобразност от реализиране на идеите за електронна община и да се организира проектирането на решенията.

С настоящото проучване целта не е да се създаде проект за електронна община, а да се наложат критериите, целите, параметрите и идеите които ще решат приоритетите при създаване на целите и задачите на проектирането на електронните услуги и движението към идеята за електронна община. На този етап задачата е да се оценят количествено тежестта на критериите за икономическа изгода и да се даде материална стойност на критериите за трудностите по реализация. Тези стойности ще залегнат в методиката като основание за оценка на приоритетността на проектите.

След тяхното формулиране ще бъдат подложени на методика за оценка за управление на портфолио от инициативи за приоритизация на проекти.

D. Методиката

Изходната точка е декомпозиране на проектното изследване, за да се формализират областите, намиране на ключовите функционалности в непосредствената работа и оценка в пространство на икономически ползи и леснота на реализация на планираната идея.

1.ПРИНЦИПИ ЗА ПРИОРИТИЗАЦИЯ НА ИНИЦИАТИВИ ЗА Е-ОБЩИНА

При декомпозиране на проектите могат да бъдат определени следните принципи за приоритизация:

1. Идентифициране на ключовите инициативи за еОбщина.
2. Декомпозиране на ключовите инициативи на ключови функционалности.
3. Изработване на карта на ключовите функционалности.
4. Идентифициране на връзките/взаимозависимостите.
5. Определяне на пътна карта на резултатите.
6. Определяне на параметрите за осъществяване на проекта (срок, хора, финанси, рискове и др.).
7. Мониторинг и контрол на изпълнението.

Източник: Адаптирано от Hartman & Sifonis netReady – Strategies for success in E-conomy, 2002

Във въпросите, които се анкетират, се търси такъв кръг от проблеми, че разрешаването им да послужи като гръбнак за решаване на много по-голямо множество проблеми и това решение устойчиво може да се прилага в бъдеще. Например - Единен регистър.

2.КЛЮЧОВИТЕ ФУНКЦИОНАЛНОСТИ

Стратегията за електронно правителство на България[30] разглежда четири основни аспекта на комуникация и услуги:

- "Администрация – Граждани" (А-Г) Съвременни Интернет и интранет WEB базирани решения, съчетани с традиционните средства за осигуряване на широк достъп, които да водят до качествени промени в условията за комуникиране и предоставяне на услуги за гражданите.
- "Администрация – Бизнес" (А-Б) Съвременни решения, които оптимизират процесите и деловите отношения между администрацията и различните икономически субекти.
- "Администрация – Администрация" (А-А) Развитие на информационните технологии в национален и междудържавен аспект с оглед на ефективно взаимодействие между различните административни структури.
- "Институционална ефективност и ефикасност" (ИЕЕ) Организиране и оптимизиране на бизнес процесите, на отношенията "Администрация – Служители" и на комуникацията в отделните административни структури.

[30] http://www.ccit.government.bg/documents/e-govstrategia.doc

Въз основа на опита на Cisco в областта на електронния бизнес и проведеното проучване на готовност за въвеждане на информационните технологии в централната държавна администрация, могат да се добавят още две области:

- “Е-култура” Организационни, лидерски и други инициативи, които са предназначени да променят културата на хората и да възстановят взаимоотношенията на държавата с гражданите и бизнеса
- “Уеб инфраструктура”

Инфраструктурата, необходима за предоставянето на всички електронни услуги и комуникации от държавата.

Това означава, че във въпросите се търси отговор за наличието на идеи за систематизираност на услугите и насоката им към персонализация, самоорганизираност и ефективно присъствие в информационната среда.

Матрицата за приоритизиране на проекти, разработена от A.Hartman и J.Sifonis осигурява прост и бърз анализ на портфолио от проекти. Тя е удобен инструмент за превръщане на списък от “желани проекти” в пътна карта за ефективно постигане на определена цел.

По ординатата на матрицата се прави оценка на икономическите ползи от осъществяването на определен проект, а по абсцисата проектът се оценява по леснота на изпълнение.

3.ПОДКРИТЕРИИ ЗА ОЦЕНКА НА ИКОНОМИЧЕСКИТЕ ПОЛЗИ.

Всички произхождат от: II. ИДЕЙНИ ЦЕЛИ В РЕАЛИЗАЦИЯТА НА eGovernment.

В тази група са изведени следните критерии:

- **Ефективност**. Възвращаемост на инвестицията, разглеждана като намаляване на разходите за извършване на услугата и/или повишаване на събираемостта на данъци и такси в сравнение с направените инвестиции и оперативните разходи необходими за поддържане на системата през периода на полезното й използване. Оценява се степента, до която инициативата е полезна за решаване на бизнес нужда по време на целия й жизнен цикъл и води до повишаване на производителността и намаляване на разходите.
- **Стратегическо съответствие**. Значимост на проекта за икономическото и социалното развитие на страната. Степента, до която проектът ускорява икономическото и социалното развитие, пести време и ресурси на гражданите и фирмите, елиминирайки по този начин административните бариери. До колко се увеличава прозрачността в държавната администрация и се намалява корупцията. Съвпадение с приоритетите на Европейския съюз оценено по степента, до която проектът е в съответствие с индикаторите за оценка на електронното правителство в e-Europe Action Plan (индикативния списък с 20 услуги в Приложение 3 на Стратегията за електронно правителство)
- **Значение**. Степента, до която инициативата посреща нуждите на широк кръг от вътрешни и външни клиенти например: брой на организации, процеси и граждани, засегнати от внедряването на проекта;
- **Устойчивост и скалируемост**. Степента, до която задачата има възможност за възпроизводство на вложените ресурси и лесно може да бъде използвана като база за бъдещо развитие.
- **Пазарни условия**. Степента, до която проектът не изисква някакви специални условия за използването на данните и предоставянето на услугата. Степента до която фирмите, служителите и/или гражданите са готови да използват услугата. До каква степен съществува търсене на дадената услуга.
- **Риск**. Степента, до която проектът не е застрашен от провал, измерван като значителни загуби на финанси, ресурси или функции.
- **Икономически ефекти за информационните и комуникационни технологии /ИКТ/ в бизнеса и обществото**. Степента до която съответната инициатива води до положителни ефекти за **ИКТ** пазара, конкурентоспособността на България, заетост на висококвалифицирани специалисти и др., положителни ефекти за бизнеса и обществото.
- **Видимост за медиите и обществото**. Степента, до която задачата ще бъде поставена във фокуса на обществено внимание и ще осигури подкрепа за изпълнението на стратегията. Степента, до която инициативата посреща нуждите на широк кръг от вътрешни и външни клиенти например: брой на организации, процеси и граждани, облагодетелствани от внедряването на проекта.

- **Международен имидж.** Степента, до която задачата допринася за повишаване имиджа на България в **ИКТ** по света.

4.ПОДКРИТЕРИИ ЗА ОЦЕНКА НА ЛЕСНОТАТА НА ИЗПЪЛНЕНИЕ

- **Ресурси/Технология**. Степента, до която инициативата е планирана по отношение на финансиране, нужни умения, необходимо обучение, ИТ стандарти, инфраструктура. До колко е достатъчна съществуващата инфраструктура и е осигурена наличност на необходимите ресурси.
- **Оперативна/организационна готовност**. Степента, до която инициативата взема предвид развитието на бизнес процесите на електронното правителство. Оперативната готовност на организацията да поддържа новите бизнес процеси в електронното правителство.
- **Взаимно допълване**. Степента, до която инициативата е взаимосвързана. Инфраструктурните, оперативните и други разходи могат да бъдат взаимно допълвани чрез комбиниране с други инициативи с подобни нужди.
- **Външни задържащи фактори**. Степента, до която законови, регулаторни механизми и/или възможностите на заинтересованите лица НЕ препятстват нормалната реализация на проекта.
- **Организационна и оперативна готовност (Култура).** Степента, до която има организационна и оперативна готовност за изпълнение на задачата и тя отговаря на организационната култура на съответните изпълнители. Културна готовност и наличие на умения за управление на промяната по отношение на новите бизнес процеси.
- **Човешки ресурси и натрупан опит**. Степента, на наличие на позитивен опит и квалифицирани човешки ресурси за изпълнение на съответната задача. (Необходимостта от допълнително обучение и квалификация намалява оценката по този показател защото забавя процеса на реализация)
- **Финансови ресурси и контрол на разходите**. Степента, до която инициативата е планирана и може да бъде обезпечена по отношение на финансиране и други ресурси. Степента, до която разходите за разработка на решението и контрола на текущите разходи са управляеми по време на процеса на изпълнение.

В резултат се онагледява следното:

- Къде върху матрицата са концентрирани проектите/инициативите?
- Какъв е бизнес фокусът на вашата инициатива?
- Какви са трите основни причини за този бизнес фокус?
- Каква е позицията на вашата инициатива сравнена с общото равнище на страната, Европа, света?

E. Процесът на анкетиране

С последващите въпроси си поставяме задача да извлечем приоритетните инициативи, като ги групираме според горните принципи и търсим проявлението в реалната практика на обслужване. Виж ПРИЛОЖЕНИЯТА.

F. Приложения

Моля, да се отговори кратко – с 1-2 изречения или думи, но писмено и да се окаже съдействие на анкетиращите на място.

Типове въпроси:

Група 1*. Как се реализира процесът на обслужването? Реална практика и оперативна готовност*

1. Кои са най-големите проблеми които общината среща при обслужването на гражданите?
2. Кои са най-често използваните услуги?
3. Колко услуги предоставяте?
4. Как се регистрират заявките за услуга или непосредственото им предоставяне?
5. Кои са типичните клиенти на услугата?
6. Как се обслужват инвалиди?
7. Как се описва резултата от предоставянето на една услуга?
8. Има ли двустранна връзка с данъчната администрация?

9. Кои услуги могат да се получават автоматично?
10. Каква група потребители касае най-много?
11. Как се взема решение за предоставяне на една услуга и какво да се иска при заявяването и?
12. Има ли персонална или отговорност свързана с дадения пост при предоставяне на услуга?
13. Има ли връзка между предоставените услуги? От една да следва, или да се изисква да е получена друга.
14. Систематизирана ли е нормативната база за типа предоставени от вас услуги?
15. Има ли комплексни и частични услуги? (Напр. Разрешение за строеж изисква скица, която също е услуга.)
16. На колко места се извършват услугите? Колко души едновременно?
17. Има ли услуги които се извършват на едно гише?
18. Има ли център за информационно обслужване на гражданите?
19. Какво е качеството на получената услуга, давността, валидността?
20. Има ли връзка между изходните и крайни информации за една услуга?

***Група 2**. Каква е информационната ви осигуреност в момента? Готовност и ресурси.*

1. Какви средства имате за събиране на информацията?
2. Имате ли технология и средства за организация и обмен на информацията? – Интернет, мрежа – докъде
3. Има ли необходимата мрежова инфраструктура?
4. Има ли необходима техническа и образователна обезпеченост на комуникациите?
5. Каква специфична информация има във всеки регистър?
6. Как се отчита общото в информацията и се организират информационните потоци?
7. Има ли единна информационна система и какви услуги обслужва?
8. Как се обменят данни с националните и други регистри и информационни системи?
9. Какви специфични информационни системи има в общината?
10. Как се реализира документооборота при вас?
11. Как се класифицира информацията при вас и организира достъпа до нея?
12. Как се предоставя информация получена по електронен път на гражданите?
13. Как практически всеки гражданин има достъп до информацията събирана за него?
14. Какви мерки за сигурност, защита и достъп до информацията са предвидени? – правят ли се резервни копия, криптира ли се предаваната информация и др.
15. Как практически е организирана хартиената информация в момента? – папки, списъци, свързана информация и др.
16. Как се индексира и поддържа информацията в актуално състояние?
17. Как се събира и предава информация в районите и от районите?
18. Как оценявате грамотността на потребителите за получаване на информация от вашите системи?
19. Колко често се налага обучение при вас?
20. Как получавате информация, необходима за работата ви?
21. Имате ли съвместни проекти с външни фирми и организации?
22. Имате ли програма и мероприятия за управление на риска от унищожаване и изгубване на информация?

***Група 3.** Организационните въпроси в общината, свързани с обслужването, нормативи и нормална практика*

1. Има ли правилник за обслужването, реда на промяна на неговото съдържание, обслужване, актуализиране и достъп?
2. Има ли правилници за функционирането на органите и службите и как действат практически?
3. Има ли финансова оценка на себестойността на предлаганите услуги?
4. Има ли бюджет за изследване с цел оптимизация на процесите за обслужване?
5. Как ще се повиши броя на обслужвани при въвеждане на електронни услуги?
6. Колко е срокът за решение на общински съвет, кмет?
7. Как се регистрират и обслужват жалби, оплаквания и др. ?
8. Как се проследява жизнения път на един такъв документ?
9. Как са определени отговорните лица в документооборота?
10. Има ли финансов план за иновационна политика и електронно правителство?
11. Какво е ролята на информационния отдел при определяне на новите услуги?
12. Имате ли достъп до електронни регистри?
13. Има ли формализирани документи за вход и изход от регистрите – напр. Бланка за заявление, за молба, удостоверение, права и доколко са еднотипни за различните услуги и служби ?
14. Какви мерки се вземат за проверка и адекватност на информацията?
15. Как се разпространява информацията до службите и районите?
16. Как се организира информацията за вътрешното функциониране на общината?

17. Как се оценява стойността на услугата?
18. Компенсира ли се извънредния труд?
19. Как се оценя себестойността на комуникациите?
20. Как се планира бюджет за информационни услуги?

Група 4. *Външни и задържащи фактори, eGovernment*

1. Кои услуги могат да се получават автоматично?
2. Има ли условия при които могат да се откажат услуги?
3. Кои услуги според вас изискват „електронен подпис"?
4. Как може за се използва удостоверението за електронен сертификат във вътрешната организация на работата ви?
5. Кои услуги изискват администриране – едноличнo решение?
6. Решаването на кои услуги изисква само подпис от упълномощено лице?
7. Какви разрешителни режими, лицензиране и др. има в общината?
8. Кои услуги се решават чрез комисии?
9. Каква информация е публикувана за услугите в общината?
10. Какви причини намирате, за да не може да се смени адресната регистрация по електронен път?
11. Има ли актуален Интернет сайт за вашите конкретни услуги?
12. Има ли осигурен бюджет за актуализацията на сайта?
13. Имате ли приемане по електронен път на заявка за обслужване?
14. Как става връщане на резултат от това?
15. Къде се регистрира процеса на обслужване и проблемите свързани с него?
16. Как се извършва мониторинг на обслужването?
17. Как се получава, въвежда и използва информацията свързана с eGovernment?
18. Имате ли готовност да издавате по електронен път свидетелство за раждане?
19. Как се организира достъпът до класифицирана информация?
20. Има ли електронно плащане за услуги?
21. Как се получава информация за специализации, квалификации и др.?
22. Имате ли финансовата, физическата и организационна възможност за развитие на Информационните центрове за граждани?
23. Как се отразяват промените в нормативната уредба в процедурите за обслужване?
24. Как се следят промени в законодателството?
25. В сегашния момент опишете какво трябва да направи гражданин, за да се отпише от един адрес и да се запише на друг в рамките на общината.

Група 5. *Как се реализира Видимост за медиите ?*

1. Как се информира медиите за решенията на ръководните органи?
2. Какъв е редът за представяне на информация в медиите?
3. Имате ли необходимата реклама на услугите ви?
4. Имате ли информационни и рекламни брошури?
5. Имате ли PR стратегия за медийно присъствие?
6. Как ще публикувате новите електронни услуги на общината?
7. Колко често журналисти получават информация от вас?
8. Правите ли рекламни кампании за услугите в общината?
9. Кои обществени мероприятия но общината се отразяват най-често в пресата?

G. Резултати

ПОДКРИТЕРИИ ЗА ОЦЕНКА ЛЕСНОТОТА НА ИЗПЪЛНЕНИЕ

ИНФОРМАЦИОННО И АДМИНИСТРАТИВНО ОБСЛУЖВАНЕ

Критерии	РЕСУРСИ ТЕХНОЛОГИЯ	ОРГАНИЗАЦИОННА ГОТОВНОСТ	ВЗАИМООБВЪРЗВАНЕ	ОПЕРАТИВНА ГОТОВНОСТ	ЗАДЪРЖАЩИ ФАКТОРИ	ЧОВЕШКИ ФАКТОР	ФИНАНСОВИ РЕСУРСИ	КАЧЕСТВО	УПРАВЛЕНИЕ И КОНТРОЛ
	Достатъчност	*Наличие на стратегия, план за развитие, цели*	*Комбинирани дейности, процеси и потоци информация*	*Формализирани документи, бизнес логика, базови информационни системи*	*Препятствия от нормативна база и административен ред*	*Квалификация, опит, обучение, ангажираност*	*Планиране, оценка, управляемост на тек. разходи*	*Система за преценка на качеството на вход /изход*	*Обратна връзка, мониторинг и контрол на процесите*
Дирекция УРОС И ЖП	Малко компютри, мрежа само в дирекцията	Нямат собствен план, стратегия за електронизация. Няма електронни регистри	Осъзнато, но неформализирано; Хартиен обмен; не описват процеси и формати	Неоптимизиран обмен и достъп; има формализирани документи; няма общи ресурси БД	Много решения с комисии; Авторитарни решения със заповед	Липсва система за оценка на участието му; Желание за обучение	Не се планират ресурси; Няма оценка на себестойността	Няма формирани критерии	Само йерархичен; няма архив на обратната връзка и анализ
С-р "Управление на нежилищни имоти"	Има компютри и мрежа; не оценяват наговарване	Движение по канален ред	По ред на деловодството; неосмислена полза от връзките	Вижда проблеми; не систематизира; не чувства процес	Намира препятствията за даденост и не реагира	Нама обратна връзка и обучение; не участва в планирането му	Няма резултат	Няма ясен резултат	Няма резултат
Сектор "Оценки"	Няма ефективна мрежа	Няма визия и не са запознати с идеите за електронно правителство	Хартиен път; няма електронни регистри; не намира последователност в ред на получаване	Оценя времето за търсене; честотата; няма електр. регистър	Много комисии;	Намат процес на получаване на информация и обр. връзка	Не са участвали в планиране	Нама критерии, но вижда истински пробеми	Регистрират обр. Връзка в хартиен вид
С-р "Оценки и разпореждане"	Не отчита мрежова структура	няма	Посочва външни връзки и взаимодействия	Типизират са вътрешни връзки; не са намерени услуги които могат да се дават автом. – напр. Изчертаване, копия	Чакат резолюции от директор, съвет	Необходимост от осъзнаване на възможностите на електр. обработка	Не се оценяват и не се планират	Не отговаря	Не отговаря
С-р "Търгове, конкурси и концесии"	1 стар компютър; мрежа само в дирекцията	Не отговаря	Отчита, не се ангажира	Няма регистри за концесии и електронни средства	Не отговаря	Не отговаря	Не отговаря	Не отговаря	Не отговаря
Сектор "Жилищна политика"	Липса на техн. обвързаност	Очаквания за централен ред	Ясни процеси но не са формализирани и алгоритизирани	Липса на разбиране за документооборот, централизирана услуга, идея за регистри	Решения с комисии; Липса на електронен документооборот	Има нужда от обвързване	Не участва във формиране на оценки	Няма отговор	Има нужда от обратна връзка
Отдел "Информационно обслужване"	Няма идеи за електронна услуга; мрежа в дирекцията	Не се осъзнава като организатор и инициатор	Само с горните нива, не са осъзнати информационни процеси	Формализирани са хартиени заявки; регистрите са хартиени; има централизирани услуги на едно гише	Не се ползват общи ресурси и бази	Има желание, но няма обвързаност	Не участва	Нямат система за обратна връзка	Контролът се замества от вертикално администриране

ПОДКРИТЕРИИ ЗА ОЦЕНКА НА ИКОНОМИЧЕСКИТЕ ПОЛЗИ

ИНФОРМАЦИОННО И АДМИНИСТРАТИВНО ОБСЛУЖВАНЕ

Крите-рии	СТРАТЕГИ-ЧЕСКО СЪОТВЕТ-СТВИЕ	ЗНАЧЕНИЕ	ЕФЕКТИВ-НОСТ	ИТ ЕФЕКТИВ-НОСТ ЗА ОБ-ЩЕСТВОТО	УСТОЙЧИ-ВОСТ И СТА-БИЛНОСТ	ПАЗАРНИ УСЛОВИЯ И ТЪРСЕНЕ	РИСК	ВИДИМОСТ И ИМИДЖ
	Значимост за икономи-ческото и социалното развитие	*За какъв кръг клиенти*	*Възвраща-емост на ин-вестицията в намаляване разходите по обслужване*	*Положителен ефект на ИТ пазара, зае-тост, за биз-неса и обще-ството*	*База за разви-тие и въз-производство на ресурси*	*Търсене на услугата и готовност на потре-бителите за ползването и*	*Степен на незастра-шеност от провал и загуба на ресурс*	*Степен на поставяне във фокуса на общест-веното вни-мание и оси-гурява под-крепа*
Дирекция УРОС И ЖП	Значителна за соц. развитие	Жилища, Конкурси, концесии	В информаци-онната обвър-заност	Антикоруп-ционна прак-тика	Прозрачност на подходите при общи бази и информаци-онни решения	КАДАСТЪР, имоти, конкурси	От необвър-заност и липса на регистри и система за контрол	Централи-зирано представяне през общи-ната
Сектор "Управление на нежилищни имоти"	За интереса на партии и организации	Процеса на представяне на инфор-мация и за експлоатаци-ята на имотите	Общуване с бизнеса и политиката – установяване на процесу-ален ред	Специа-лизиран за партиите	Единна регист-рация и крите-рии на наем-ните договори	Търсят се незаети имоти или такива, на които изтича договорът	Липса на интегритет на органите, които обслужват договорите	Би могло да има огромно внимание и да се демон-стрира проз-рачност
Сектор "Оценки"	За добра практика и вътрешен ред	Масов харак-тер на услу-гата	Възможност за автоматични копия, скици и др.	За форма-лизиране на оценките	При общи бази и инфо-ресурси	Оценка на имот	Работни комисии и експертни оценки	Имат собствена брошура за услугите в този им вид
Сектор "Оценки и разпореждане"	За добра практика и вътрешен ред	Налагане на добра практика	Възможност за автоматизация на формални справки	За формализиран е на оценките	При общи бази и инфо-ресурси	Оценка на имот и състо-янието му при разпореждане с него	От налагане на необективни оценки	Осмисляне на информация та и начина на представяне то и
Сектор "Търгове, конкурси и	Обществени поръчки	Най-широка общест-веност	Няма виждане за система	За организиране на прозрачни търгове	Обективиране на критериите	Самите услуги се търсят	От авторитарна намеса	Неосъзната необходимос т от прозрачност
Сектор "Жилищна политика"	Социална политика	Най-видим проблем	При описание на формален ред с демон-страция на добра практика	Заетост на оценителите на жилищни нужди	Демонстрация на добра практика	Резултатите от работата	Решения на комисии	Максимална демон-страция на обективност
Отдел "Информацион-но обслужване"	За вътреш-ната роля на осъзнаване на процесите и връзка между адми-нистрациите	Консоли-дираща роля и възпитание за самата администрац ия	Ако са изградени базови системи и те обслужват основата на администра-цията	Множество съвместни проекти могат да се инициират, напр. КАДАСТЪР	Липсва идея и водеща роля за основни базови информа-ционни и комуникационн и структури	Обучение, добри примери и внедряване на базови системи и технологии	Адекватност и положителн а нагласа	Нямат собст-вена политика и нагласа

Литература

Стратегия за модернизиране на държавната администрация - от присъединяване към интегриране. Приета с Решение на Министерския съвет №465 от 09.07.2002г.

Стратегия за електронно правителство. Приета с Решение на Министерския съвет №866 от 28.12. 2002г.

Анализ на действащото законодателство за възможностите, които предоставя за обмен на информация в електронен вид във връзка с приетата с решение на Министерския съвет № 866 от 28.12.2002г Стратегия за електронно правителство:

1. Общественото мнение за предоставянето на административни услуги в България - юни 2002 г.
2. Техники за управление на портфолио от инициативи Матрица на Хартман и Сифонис за приоритизация на проекти
3. Изследване за отношението на бизнеса към административните услуги предлагани по електронен път от Координационен център за информационни, комуникационни и управленски технологии към Министерския съвет и ПРООН -www.ccit.government.bg

Концепция за подобряване на административното обслужване в контекста на принципа "едно гише" Приета с Решение на Министерския съвет № 878 от 29.12.2002 г.

БАИТ и Електронното правителство в България – готовност и очаквания – ноември 2002 г.

Решение на Министерския съвет №867 от 28.12.2002г за участие на Република България в програмата Европейската общност IDA II - обмен на данни между администрациите (ОДА), за периода 2003 - 2004г.

Електронното правителство и Европейския съюз, The Internet and the City Conference "Местното електронно правителство в информационното общество" Барселона, 21.03.2003 - Speech of Mr Erkki Liikanen, EU Commissioner responsible for Enterprise and the Information Society

Implementing the President's Management Agenda for E-Government, E-Government Strategy, April 2003 http://www.whitehouse.gov/omb/egov/2003egov_strat.pdf

Local E-Government - Office of the Legislative Auditor State of Minnesota

e-Government Interoperability Framework / Version 4.0 / 25th April 2002 - The e-Envoy, Stockley House, 130 Wilton Road, London, SW1V 1LQ, ukgovtalk@e-envoy.gsi.gov.uk

eGovernment Resource Book - Synopses of IST projects relating to eGovernment - By the Commission of the European Communities, Information Society Directorate-General.

Изучение информированности и отношения населения к проектам и мероприятиям федеральной целевой программы "Электронная Россия".

Local e-government now: a worldwide view - This report was published in September 2002 jointly by the Improvement & Development Agency (IDeA) and the Society of Information Technology Management (Socitm), ISBN 1 899 330 89 5 Copyright © 2002 Socitm Services Ltd and IDeA

Информация за авторите

Димитър Любенов Николов – мениджър "Инфоком", България, e-mail: **d.nikolov@infocom.bg**

Румяна Божидарова Стефанова – Нов Български Университет – ст. асистент, ул. Монтевидео 21, София 1618, България, e-mail: **rbstefanova@abv.bg**

Ивелин Димитров Стефанов - началник отдел в "Информационно Обслужване" АД, България, e-mail: **istefanov@abv.bg**

ИНФОРМАЦИОННО МОДЕЛИРАНЕ НА ДВУКООРДИНАТНО ИЗМЕРВАНЕ НА ОПТИЧНИ ПАРАМЕТРИ

Георги Стоилов, Петър Шарланджиев, Никола Мечкаров

***Резюме**: Предложен е метод за измерване и визуализация на комплексния коефициент на пропускане (КП) за 2D микро-обекти. Методът се базира на изчисляване КП от дифракционната картина и функцията на осветяващата апертура за монохроматична светлина. За определяне на фазата на дифрактиралата светлина е използван фазово-стъпков метод.*

***Ключови думи**: микроскопия, фазово-стъпков метод, интерферометрия, обратна задача в оптиката*

Въведение

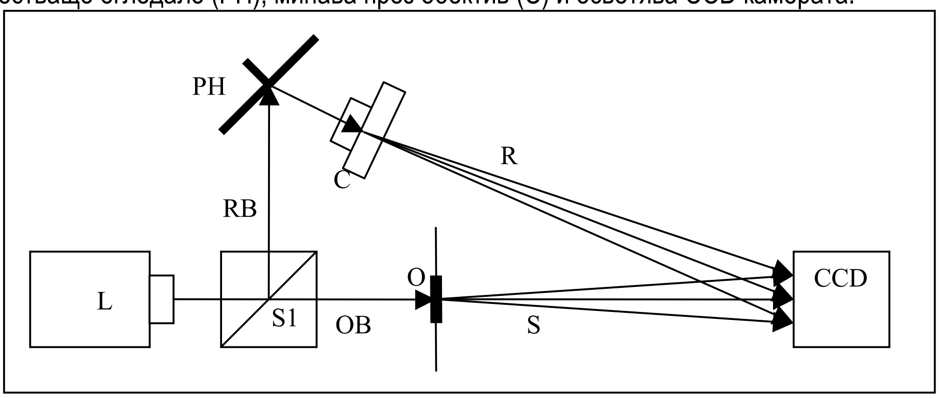
Оптичните характеристики на различни обекти са индикатор за състоянието им (клетки), качество на производство (оптични елементи), хомогенност (разтвори и гелове) и т. н. За измерване на тези характеристики се използват разнообразни лабораторни и промишлени измервателни системи. Най-общите съставни елементи са светлинен излъчвател оптична система за формиране на изображение (огледала, обективи, делители и др.) и фоторегистриращо устройство. Използването на обектив в измервателната система понякога е нежелателно, по-скъпо или дори невъзможно. Това е причината за търсене на измервателни методи на оптичните параметри на микро-обекти чрез информация, измерена във фраунхоферовата зона.

Принцип на измерване

Принципът на измерване се базира на обратната задача в оптиката [1] - по известна функция на разпределение на интензитета по апертурата на осветяващия сноп и измереното разпределение на интензитета на екран зад обекта се изчислява предавателната функция на обекта. Известно е, че използването на фазово-стъпковия метод [2] позволява измерване на комплексната амплитуда на светлината. Едновременното използване на фазово-стъпковия метод и измерването на интензитета на дифрактиралата светлина позволява намиране на коефициента на пропускане в комплексен вид. За елиминиране на неизвестните параметри на осветяващата система се използва информация от допълнителни опорни измервания.

Постановка

Една възможна интерферометрична схема е показана на фиг.1. Светлинен сноп от лазер (L) се разделя от светоделител (BS) на обектен сноп (OB) и опорен сноп (RB). Обектният сноп минава през измервания обект (O) и пропуснатата светлина се регистрира от CCD камера (CCD). Опорният сноп, отразен от фазово-отместващо огледало (PH), минава през обектив (C) и осветява CCD камерата.



Фиг. 1

Приближения и сфера на приложение

В това изследване 3D обектите се апроксимират с 2D, като предполагаме, че обектите са тънки. С тази апроксимация можем да постигнем точно едно решение. В реални 3D измервания би могъл да се използва томографски метод, като се регистрира набор от изображения при преместването на CCD сензора по направление към обекта.

Разглеждат се прозрачни изотропни обекти при които разсейването и отражението като оптични параметри на обекта не са отчетени. Има голям клас обекти, при измерването на които тези параметри също могат да се игнорират.

Използването на светлина с дължина на вълната 0.5 –1.0 μm е ограничено от характеристиките на CCD сензора. Използването на по-малка дължина на вълната води до нарастване точността на измерването според теорията на Абе.

Обикновено CCD сензорът има размер 5 – 10 mm. Това ограничава областта на използваните пространствени честоти. Когато осветяващият сноп има апертура около 5 μm и амплитудата и фазата са константни по цялото сечение, пространствените честоти са локализирани в ъгъл около 5 градуса, но ако апертурата е периодична решетка с период 1 μm, ъгълът нараства до 45 градуса. За да се използват всички пространствени честоти, камерата трябва да се приближи до обекта, което може да ни изведе извън далечната (Фраунхоферовата) зона.

Теоретичното изследване на грешките от пространственото разположение на CCD сензора, точността на АЦП, пикселизацията, дебелината на обекта и неговите оптични характеристики са важна част от подготовката на експериментална постановка. Постигане на някои от условията може да се окаже пречка на съвременния етап на развитие на елементната база.

Математически модел

Когато разстоянието от обекта до регистриращата повърхност е много по-голямо от размерите на обекта, може да се направи апроксимация на регистрираната картина с преобразувание на Фурие от изображението на обекта [3,4]. В този случай общият интензитет на светлината от обекта и от опорния сноп, измерен със CCD камерата, може да се представи с:

$$I = [F(O.S) + R]^2 \quad (1)$$

където *I* е интензитетът на всеки пиксел, *O* е предавателната функция на обекта, *S* е функцията на разпределение на амплитудата в апертурата на снопа, *R* е функцията на разпределение на амплитудата в апертурата на опорния сноп и *F* е преобразование на Фурие.

След разкриването на скобите:

$$I = [F(O.S)]^2 + R^2 + 2F(O.S).R \quad (2)$$

От (2) се вижда, че първото и второто събираеми изразяват интензитета от обект без опора и от опора без обект. За да се елиминират тези части, трябва да се направят 2 отделни измервания (два кадъра):

$$O2 = [F(O.S)]^2 \quad (3)$$

и

$$R2 = R^2 \quad (4)$$

За всеки кадър *I* изчисляваме съответното коригирано *Ic*, където:

$$Ic = (I - O2 - R2)/2 = F(O.S).R \quad (5)$$

Използвайки фазово отместване (4- стъпков метод) на *RB* чрез PH, получаваме 4 изображения:

$$Ic_0 = F(O.S).R.COS(\varphi + 0.\pi)$$
$$Ic_1 = F(O.S).R.COS(\varphi + 90.\pi)$$
$$Ic_2 = F(O.S).R.COS(\varphi + 180.\pi)$$
$$Ic_3 = F(O.S).R.COS(\varphi + 270.\pi) \qquad (6)$$

където φ е фазата на вълната, а $k\pi$ внесената фаза на опорния сноп. От системата уравнения може да бъде изчислена комплексната стойност на *Ic (изображението A4)*, което в честотното пространство е трансформацията на Фурие на апертурата на съответната осветяваща система.

$$O.S = F^{-1}(A4) \qquad (7)$$

На края трябва да се извърши опорно измерване за елиминиране на *S*. То представлява измерване без обект и физическата му същност е измерване на апаратната функция на измервателната система. По този начин се компенсира влиянието на много неточности при задаване на параметрите на измервателната система като разпределение на интензитет на осветяващия сноп, геометрия на измервателния отвор, опорен сноп и неговите параметри.

$$S = F^{-1}(Ar4) \qquad (8)$$

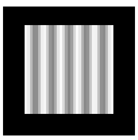
$$O = F^{-1}(A4)/S \qquad (9)$$

O е матрица от комплексни числа. Модулът изразява затихването, а аргументът – фазовото закъснение за всяка точка от измервания обект.

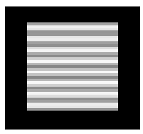
Информационен модел

Написана е програма за симулация на постановката от фиг. 1 и обработка на информацията измерена от CCD камерата. Тя се използва за оценка на изискванията към постановката и ограниченията при използване на метода. Програмата е написана на C++ за PC. В някои от случаите се използва алгоритъм за FFT, но е разработен и вариант с директно интегриране, поради факта, че не може всеки път да се удовлетворят изискванията за прилагането на FFT.

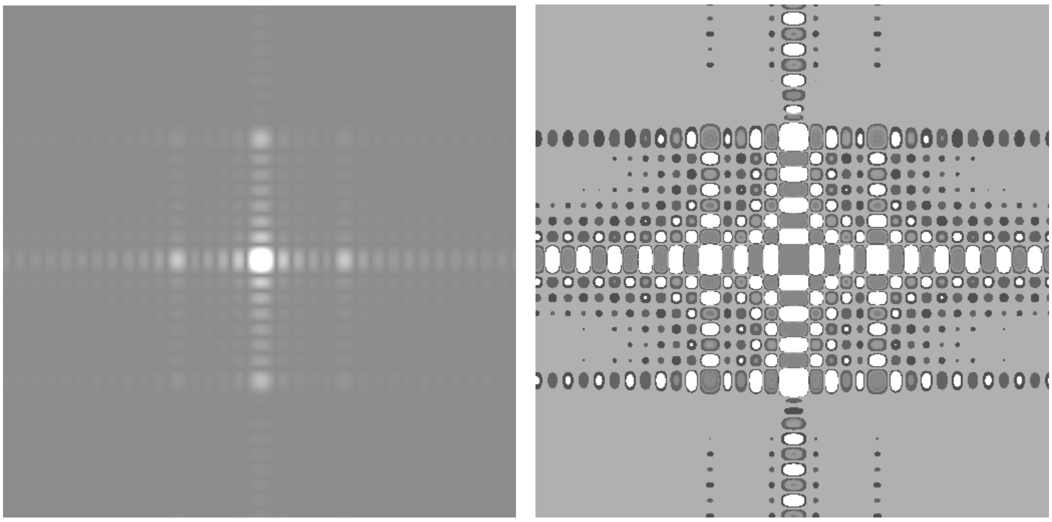
Като обект се избира амплитудна (и/или фазова) пластинка с предавателна функция O = k0 + k1 SIN (k2 X). Снимка на амплитудна пластинка е показана на фиг. 2. Като функция на разпределение на опорния сноп е избрана S = n0 + n1 SIN (n2 Y). Разпределението на амплитудата е показано на фиг. 3.



Фиг.2 Измерван обект
формула за симулация O(x, y) = 1+0.3 SIN (kX)



Фиг.3 Апертура на измервателната система
формула за симулация S(x,y) = 1+0.3 SIN(nY)

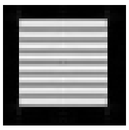


Фиг.4 Крайна информация преди обратно преобразувание на Фурие: амплитудна и фазова карта

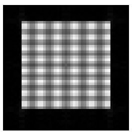
Изборът е направен, за да се види интерференцията и взаимното влияние на апаратната функция и на обектния сноп. Стойността на пикселите във всеки изчислен образ е закръглена до 0,5 % от максимума. Това е направено с цел симулация на най-популярните 8-битови фрейм-грабери.

Фиг. 4 показва амплитудна и фазова карта на изчислената функция на обекта с преобразование на Фурие.

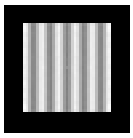
След прилагане на обратно преобразувание на Фурие се получават данни, показани на Фиг. 5, 6, 7, близки до изходните.



Фиг. 5 Изчислено разпределение на предавателната функция на системата

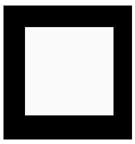


Фиг. 6 Изчислено разпределение обект + система



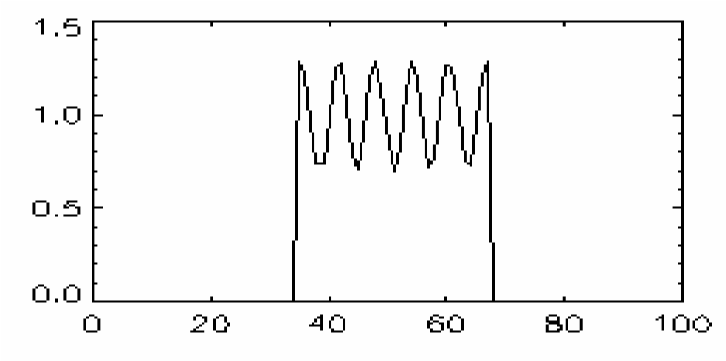
Фиг. 7 Изчислен обект

При обратното пресмятане поради закръглението на стойностите на някои пиксели функцията на обекта получава много големи стойности – над 50 пъти от максимума на измерената стойност. Това изисква използването на функция-прозорец за маскиране нежеланите и неочаквани резултати. Маскиращата функция (Фиг. 8) се базира на изчисленото разпределение обект+система. Областта със стойности над предварително зададен праг се разглежда като маска. За елиминиране на малки единични петна се използва изглаждащ филтър с подходяща апертура (2 – 10 пиксела).



Фиг. 8 Изчислен прозорец за валидна информация

Фиг. 9 показва разпределението на коефициента на пропускане в средата на изображението (в напречно сечение), който в случая е симулиран със стойности даже по-големи от 1.



Фиг. 9 Напречно сечение на интензитета в средата на възстановен образ

Анализ и заключение

Измерването на обекти по-големи от 5 – 10 μm води до позициониране на камерата на повече от 0.1 – 0.2 m от обекта. Приложението на метода в този случай е неподходящо и дори невъзможно. От друга страна наблюдението на малки обекти води до нарастване на високочестотните компоненти в пространствените честоти. Регистрирането на тези спектрални компоненти изисква CCD сензор с голяма повърхност.

Резултатите от моделирането ни позволяват да твърдим, че за визуализация на фазови обекти е възможно използването на 8-битов АЦП. В случай на използване на тази техника за измерване се изисква АЦП с по-голяма разрешаваща способност. Мощността на лазера и чувствителността на CCD сензора са съществени, защото поради високата енергийна плътност обектът ще бъде повреден. В предложения модел е наложено и ограничението обектът да е изотропен. Тези ограничения ни дават основание да предположим, че подходът за измерване на оптичните характеристики да двумерни обекти с използване на решаване на обратната задача в оптиката комбиниран с фазово-стъпковия метод е особено подходящ за наблюдаване на клетки. Наблюдаването на живи клетки в течна среда би могло да се осъществи с използването на импулсен лазер.

Библиография

1. Baltes H. P., Inverse Source Problems in Optics, Springer-Verlag Berlin-New York, 1978.
2. Tiziani H.J., Optical methods for precision measurements, Optical and Quantum Electronics 21 (1989) 253-282.
3. Cowley J. M., Diffraction Physics, North-Holland Publishing Company, New York, 1975.
4. Rastogi P., Sharma A., Systematic approach to image formation in digital holography. Opt. Eng. 42(5) 1208-1214 (May 2003).

Информация за авторите

Георги Стоилов – ЦЛОЗОИ, научен сътрудник, gstoilov@optics.bas.bg

Петър Шарланджиев – ЦЛОЗОИ, старши научен сътрудник, pete@optics.bas.bg

Никола Мечкаров – ЦЛОЗОИ, научен сътрудник, nikola@optics.bas.bg

ЦЛОЗОИ – Българска Академия на Науките, София 1113, ул. „Акад. Г. Бончев" 101, П.К. 95